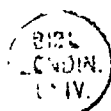


Thesis title: The phonology of present-day Cantonese

Name of candidate: Cheung, Kwan-hin

Degree submitted for: PhD 1986

College: University College



A B S T R A C T

This thesis describes the phonology of present-day Cantonese. In addition to tone, onset and rime, the thesis also covers realization, variation, casual speech and intonation. A separate chapter considers the syllable as a whole. With sympathetic understanding, the thesis reviews previous work on the subject. In doing so, it tries to provide principled answers to the questions how and why Cantonese phonologies differ. In its own treatment of the subject, it benefits from indigenous Chinese phonology, classical phonemics, Firthian prosodic phonology, SPE phonology, and autosegmental phonology, as well as European structuralism, while dismissing the time-honoured principle of unilinear phoneme-size segmentation as inappropriate for Cantonese. The mora is introduced into the organization of Cantonese sounds. The descriptive device of autosegmental phonology enables us to consider morae as "autosegments", thereby capturing a number of regularities which are otherwise difficult to characterize elegantly. Another innovation in the thesis is the idea of "coercion", a process whereby uncanonical phonetic forms, which arise as the output of casual speech processes, are replaced by canonical forms. The mora, coercion, and autosegmental representations together account for a good deal of lower-level regularities, especially in casual, connected speech. They also contribute to understanding the discrepancies among different phonologies of Cantonese. By enabling a dynamic and holistic view of the organization of Cantonese sounds, they cast light on the static and fragmentary nature of many prevailing views on the subject.

T A B L E O F C O N T E N T S

Abstract of thesis

Table of contents

Acknowledgments

CHAPTER 1: INTRODUCTION

1.1 Aim and scope

1.2 Presentation of thesis

1.2.1 Strategy of presentation

1.2.2 Organization of thesis

1.3 Previous work on the subject

1.3.1 Why works differ?

1.3.2 Survey of literature

1.4 Framework of description

CHAPTER 2: THE REFERENCE DESCRIPTION

2.1 Presenting a standard description

2.2 Departures from Standard Description

2.2.1 Omissions in Standard Description

2.2.2 Over-differentiations in Standard Description

2.2.2.1 n- vs l-

2.2.2.2 ŋ- vs Ø-

2.3 Summary

CHAPTER 3: TONE

3.1 The inventory of tones

3.1.1 Occlusive tones

3.1.2 Suggested merger of T3 and T5

3.1.3 Suggested split of T1

3.2 Tone modification and related questions

3.2.1 The high-rising modified tone and T2

3.2.2 The high-even modified tone and T1

3.3 The characterization of tones

FRONT

CHAPTER 4: RIME

4.1 The characterization of codas

4.2 The treatment of vowels

4.2.1 Adjustments to Reference Description

4.2.1.1 The treatment of non-low short vowels

4.2.1.2 The treatment of y:

4.2.1.3 Resultant arrangement of vowels

4.2.2 A critical account of other treatments of vowels

4.2.3 The characterization of vowels

4.3 Characterizing the wellformed rime

CHAPTER 5: ONSET

5.1 Clustering

5.1.1 Setting the background

5.1.2 The "zero initial"

5.1.3 gw- and kw-

5.1.4 Consonant-lateral clusters

5.1.5 Summary

5.2 The characterization of onsets

5.2.1 Place of articulation

5.2.2 Manner of articulation

5.3 Comparison with coda

CHAPTER 6: THE MORA

6.1 Vowel-coda length complementarity

6.2 A moraic interpretation of vowel and coda length

6.3 The mora and syllable isochrony

6.4 A moraic characterization of rime types

6.5 Mora versus feature

6.6 The place of the mora in the syllable

CHAPTER 7: THE SYLLABLE

7.1 Syllabic nasals

7.2 Syllable-constituent combination restrictions

7.2.1 [*occlusion]

7.2.1.1 With tone

7.2.1.2 With rime

7.2.2 Onset

7.2.2.1 With tone

7.2.2.1.1 Nasals

7.2.2.1.2 Stops

7.2.2.2 With coda

7.2.2.3 With rime

7.3 Characterizing the wellformed syllable

7.3.1 The syllable structure revisited

7.3.2 Summarizing the wellformed syllable

CHAPTER 8: REALIZATION

8.1 [*occlusion]

8.2 Tones

8.3 Vowels

8.4 Codas

8.5 Onsets

8.6 The place of articulation of sibilants

8.7 Lip-rounding harmony

CHAPTER 9: VARIATION

9.1 A taxonomy of pronunciation variation

9.2 Systemic variations

9.2.1 [* merger of n- and l-]

9.2.2 [* merger of ŋ- and Ø-]

9.2.3 [* merger of -ŋ and -n]

9.2.4 [* merger of T3 and T5]

9.2.5 [* split of T1]

9.3 Combinational variations

- 9.3.1 Vowel with -ŋ
- 9.3.2 ŋ/θ- with tone
- 9.3.3 gw/kw- with vowel
- 9.3.4 Sibilant onsets with vowel
- 9.3.5 æ:/iɔ with coda
- 9.3.6 [ŋ] ~ [m]

9.4 Lexical-incidental variations

- 9.4.1 Rules and the lexicon in pronunciation variation
- 9.4.2 Semi-regular lexical-incidental variations
- 9.4.3 Merger-suggestive lexical-incidental variations
- 9.4.4 Isolated lexical-incidental variations

9.5 Correlating pronunciation variables with non-linguistic variables

- 9.5.1 Geographical variables
- 9.5.2 Chronological variables
- 9.5.3 Social variables
- 9.5.4 Stylistic variables

CHAPTER 10: BEYOND THE SYLLABLE

10.1 Casual speech

- 10.1.1 Cross-syllable assimilation
 - 10.1.1.1 Contiguous assimilation
 - 10.1.1.2 Non-contiguous assimilation
- 10.1.2 Non-initial onset lenition
- 10.1.3 Contraction
 - 10.1.3.1 Mora deletion
 - 10.1.3.2 Coda deletion
 - 10.1.3.3 Cl- formation
 - 10.1.3.4 Bisyllabic fusion

10.2 Intonation

- 10.2.1 Intonation in the narrow sense
- 10.2.2 Modification of the final tone

REFERENCES

A C K N O W L E D G M E N T S

I wish to thank D.C. Bennett, J Harris and A Mtenje for commenting on certain points contained in earlier versions of the thesis; P Davis, Jiāng Yùxīng, R McKee, I Roca and H S-H Yeung for providing me with very useful material which is otherwise not available to me; J Y-M Chan, J Crompton and Qǔ Lěilěi for helping me in technical matters in final-stage thesis production; and the Department of Phonetics and Linguistics, UCL, for allowing me to use the word-processing facilities of the Department.

No words can express my gratitude to J.C. Wells, my supervisor, for the amount of hard work in improving various drafts of the thesis, for his encouragement and for his unfailing and literally astonishing efficiency.

CHAPTER 1: INTRODUCTION

1.1 Aim and scope

This thesis describes the phonology of present-day Cantonese.

By description I mean that the study follows the not very long but by now firmly established tradition of descriptive linguistics. Chomskyan linguists distinguish between observational, descriptive and explanatory adequacy. I find the distinction between observational and descriptive adequacy a convincing and useful one; needless to say I shall not be content with my account being merely observationally adequate. The distinction between descriptive and explanatory adequacy, however, is not crystal clear to me. The distinction sounds straightforward; thus Chomsky (1965:25) writes:

To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of explanatory adequacy.

But such selection presupposes an established general theory of grammar, with built-in evaluation measures and a flawless grasp of language universals. Such presupposition is in my opinion premature. Stich (1972:141) points out an obvious difficulty involved:

In constructing an acquisition model, the first few plausible (approximations of) descriptively adequate grammars [*dags*] have a profound influence. For it is the abstract features of these grammars which are taken as quasi-universals. Yet the selection of these first *dags* over indefinitely many alternatives is completely unmotivated by any linguistic evidence. Which *dag* is first constructed is largely a matter of historical accident. But the accident casts its shadow over all future work. The acquisition model serves to direct future research into the channel forged by these first grammars, even though there are indefinitely many other possible channels available.

On the other hand, it can be argued that a good deal of sensible evaluation of grammar can be and has been conducted in accordance

with the requirements of descriptive adequacy. Thus, rather than pretend to explain, I am content to describe in what I think to be the most adequate way.

"Phonology" is construed to be, at one end, sufficiently different from morphology in principle, and at the other, wide enough to include a fair amount of phonetic detail. The demarcation between phonology and morphology is a matter of some controversy. The indeterminacy lies in which side to place morphophonological alternation. This thesis takes the narrower view of phonology in this respect. Morphology is referred to only if it sheds light on phonology proper; whether this is labelled "morphophonology" is a separate question. Since there is very little morphophonological alternation in Cantonese, and even this small amount is subject to lexicalization, there is no significant consequence of adopting either view of phonology.

At the more concrete end, the phonetic commitment of the thesis means that problems of variation will not be dodged. And quite independently from this reason, I hold that pronunciation variation is an integral part of the phonology of any language. An important reason for this is that we are considering a speech community, not an idiolect. Moreover, variation also applies to individuals.

Cantonese is the standard variety among the Yuè¹ dialects,² which in turn constitute one of six or seven Chinese dialect groups.³ I make a distinction between what I call "mainstream" Cantonese, spoken by a total of some ten million people in Hong Kong, Guǎngzhōu and Macao, and other regional varieties of Cantonese. It is mainstream Cantonese which is the chief object of description in the thesis. When we deal with regional variation in Section 9.5.1, however, we shall

¹ Transliteration of Chinese words other than for the illustration of Cantonese is done in the *Pinyin* system official in the People's Republic of China. As the tone marks render them sufficiently different from English orthography, the transliterated words will not be italicized from now on.

² 'Cantonese' is used by some writers to refer to the Yuè dialects collectively.

³ The other six are Northern (including Mandarin), Wú (including Shànghǎi), Mǐn (incorporating Northern, Southern and Eastern Mǐn), Xiāng, Hakka and Gàn. The last two may be conflated. (Zhān 1981)

briefly describe one other variety, namely Malayan Cantonese, which will also be sporadically referred to elsewhere when the need arises.

The attribute "present-day" in the title cannot be over-emphasized, for much of the inadequacy in previous work stems from confusion between diachrony and synchrony, between etymology and phonology, and between obsolete and current forms. The attribute "synchronic" could have been used instead. I avoid it because (i) the thesis includes a section (9.5.2) on chronological variation and (ii) I do go into diachronic description at certain points, if only to show more clearly the distinction between what is and what used to be, and to understand why something appears to be or is thought to be.

1.2 Presentation of thesis

In this section I first explain my strategy of presentation and then I outline the organization of the thesis.

1.2.1 Strategy of presentation

As Cantonese phonology is no virgin territory, familiarity with scholarship in this area of study is of the utmost importance. Though I may disagree with previous work on particular points, other writers' major contributions must be acknowledged and their main views represented and evaluated. Accordingly, the thesis abounds with references to previous work, in the form of (a) direct quotations, (b) paraphrases, (c) summaries, (d) interpretations and re-interpretations, (e) inter-analysis comparisons, (f) comments, (g) evaluations.

The first three, i.e. (a) to (c), are relatively straightforward tasks. As we know, observational adequacy is a necessary condition for descriptive adequacy. I am therefore at pains to establish the occurring forms which other writers tend to ignore or are not aware of. I am a native speaker of Cantonese. Though for certain scholarly practices in linguistics this means that all that is needed is for me to claim the grammaticality of a certain linguistic form, I proceed with more caution than that. Thus, I find support from other writers as best I can, so as to ensure that I am not deceiving the rest of world or myself. Besides,

when it comes to plain phonetic description, which relies on physical measurement or acute perception rather than argumentation, I also make reference to acoustic studies or the judgment of reliable phoneticians (notably Daniel Jones and Yuen Ren Chao) for the support of my description, though I reserve the right to challenge their judgment. For references for the above purposes, I usually do not need to go beyond (a), (b) and (c), which, as I said, are relatively straightforward tasks.

The last four, i.e. (d) to (g), which are necessary for other purposes, are not that straightforward on the other hand. They are impossible or dangerous without a thorough, sympathetic understanding of individual studies and a good grasp of previous work in general. It is for this reason that despite frequent reference to previous work throughout the thesis, I devote a separate section (1.3) to a principled overview of previous work.

A considerable part of the thesis is concerned with agreeing and disagreeing with other writers. Very often this means argumentation. Argumentation is also crucial when I am exploring new ideas (for example the idea of autosegmental morae) or new areas (for example casual speech phenomena) in Cantonese phonology.

Exposition is another mode of presentation in the thesis. It is needed for the representation of other writers' views and judgments, for the display of primary linguistic data, and in areas where systematic and thoughtful presentation is more important than arguing, such as when we deal with realization (Chapter 8) and variation (Chapter 9). For exposition I strive for (i) logical, principled, well-motivated taxonomy (e.g. making sure that classifications are mutually exclusive and collectively exhaustive) and (ii) clarity and compendiousness of presentation. In connection with (ii), I do not hesitate to use graphical representations, mainly in the form of tabulation. In connection with both (i) and (ii) the idea of binary distinctive feature matrices in modern phonology epitomizes a superb system of global classification (which lends itself particularly well to tabulation) where a set of (ideally binary) parameters, each representing a dimension in its own right, cross-classifies a set of entities. Thus more than once I borrow not only the idea but also the format of distinctive features for

the elucidation of a complex situation involving a number of cross-cutting factors.

Cantonese is not for writing down; when people do write it down Chinese characters are used. Unlike for Mandarin, no romanization system for Cantonese can claim to be standard or representative. Broad phonetic/phonemic transcription (which does not necessarily reflect the most adequate analysis of Cantonese sounds) is used here for the representation of both lexical items and sounds in Cantonese. Except for the representation of tone, for which a numerical superscript (from 1 to 6) is attached to the end of a syllable, the notations are based on the International Phonetic Alphabet (revised to 1979), subject to slight deviation from it for typographical or theoretical reasons.¹ The notations are subject to minor revision as the thesis develops, reflecting revision in the judgment or analysis of Cantonese sounds. Solidi (//) are avoided because (i) the transcription is not strictly segmental-phonemic owing to the prosodic slant of the thesis; (ii) often whether a difference is contrastive or not is uncertain, is under discussion, or is variable; and (iii) the transcriptions are easily enough distinguishable from ordinary English orthography. Narrower transcriptions are used if necessary, as when we deal with realization and variation. Square brackets ([]) are used to mark them only when confusion might arise.

Other strategic considerations will be mentioned when we consider the organization of the thesis in the next section.

1.2.2 Organization of thesis

The layout of the thesis should be clear from Table of Contents, where we find the heading of each section and subsection. This section gives supplementary information on the organization of the thesis.

The introductory chapter is preliminary to the thesis proper, which starts from Chapter 2. But it is essential for the appreciation of the entire thesis.

¹ 'g', for examples, is used in place of 'g'.

Chapter 2 develops a "reference description" of Cantonese sounds, which is the result of a "standard description" undergoing certain adjustments. The standard description represents the widely accepted core in Cantonese phonology. The adjustments are made so as to fulfil the requirement of observational adequacy. The chapter is strategically placed here so that we can use this relatively uncontroversial core as a point of departure for discussions in the rest of the thesis.

Chapters 3 to 5 deal with tone, rime and onset respectively, representing the three traditional divisions of the Cantonese syllable. To write a phonology of Cantonese, I could have stopped at the end of Chapter 5. But Chapter 6 embodies a major innovation of this thesis: the mora is introduced, and it is considered to be autosegmental. The chapter presents a self-contained justification of the autosegmental mora, but the idea also contributes to the descriptions in Chapters 7 and 10. Chapter 7 considers the syllable as a whole, relating tone, rime, onset, and [±occlusion] (which is extracted from tone and/or rime) one to another and to the syllable as a whole. The chapter is important for an appreciation of the syllable as the primary (i.e. most important, not smallest) isolate for the description of Cantonese sounds.

The last three chapters deal with topics that are often ignored or neglected by Cantonese phonologists. Chapter 8 presents the realizational regularities and details of the more or less abstract phonological entities. Chapter 9 is a comprehensive treatment of variation. Chapter 10 goes beyond the monosyllable to consider casual speech and intonational phenomena.

1.3 Previous work on the subject

We mentioned in Section 1.2 the need for "a sympathetic understanding" of previous work. "Sympathetic understanding" means that not only do we have an undistorted picture of the claims, ideas and arguments of individual works, but we also understand how those claims, ideas and arguments have been shaped by the writer's theoretical background and the objectives of his study. Accordingly Section 1.3.1 considers the principal factors that determine the way one

does Cantonese phonology. Section 1.3.2 looks at individual works, which are often better understood in the light of these factors.

1.3.1 Why works differ?

Six parameters can be isolated which contribute to shaping the way one describes the sounds of Cantonese. These are explained one by one and their implications discussed.

[+conservative]. This refers to the extent to which a writer recognizes innovations in the language. It is a shame if a (descriptive) linguist does not pay attention to such innovations. In the study of Cantonese phonology, conservatism prevails. This accounts for the persistent omission of the rimes $\epsilon:w$, $\epsilon:m/p$, $\epsilon:n/t$ and om/p . Conservatism sometimes deteriorates into prescriptivism. Thus, emerging forms are branded "wrong", so as to preserve the validity of the orthodox characterizations of the sound pattern; facts of pronunciation variation are played down or dismissed. It is interesting to note that while Jones and Woo 1912 and Chao 1947 are basically [-conservative], as can be seen from the former's recognition of the high even variant of tone 1 and the rime $\epsilon:w$, and the latter's firsthand reporting of the variations (in the system of onsets) $n- \sim l-$ and $\eta- \sim \emptyset-$, their works are so out-of-date that their descriptions may look conservative nowadays. It is a pity that more recent works scarcely inherit the non-conservative outlook of these early writers on Cantonese sounds.

[+dialectology]. This refers to whether Cantonese is described in the context of describing some other Chinese dialect or studied in its own right. [+dialectology] implies the following:

- (1) [+indigenous] (See below).
- (2) "Occlusive tones" (See Section 3.1.1) are recognized, which in turn means that at least nine tones are recognized.
- (3) A "medial" $-w-$ is recognized (See Section 5.1).

The reason for (1) is that the indigenous Chinese phonological framework is much better suited than others to Chinese interdialectal comparison. By the same token, occlusive tones and the medial $-w-$, though inappropriate for present-day Cantonese even if a basically [+indigenous] position is adopted, are nevertheless extremely useful for

interdialectal comparison.

[±indigenous]. This refers to whether the descriptive frame of indigenous Chinese phonology is adopted. [+indigenous] implies the following:

- (1) [+phonemic] (See below).
- (2) The syllable as a primary isolate is recognized.
- (3) The tripartite division of the syllable into tone, initial/onset and final/rime.
- (4) Rime = Vowel + Coda, where Coda may be vocalic.

To be consistent, occlusive tones should be recognized and -p, -t, -k treated as "allophones" of -m, -n, -ŋ. However, influenced by classical phonemics, Cantonese phonologists (including those that are [+indigenous]) tend to be reluctant to attach greater importance to tone than to segments. As p, t, k are contrastive with m, n/l, ŋ respectively as onsets they are also regarded as so contrastive as codas, while the occlusive tones are either dismissed or still maintained redundantly (See Chapter 3 for details).

[±phonemic]. This refers to whether sounds in language are presumed to be organized in the form of a single, non-hierarchical string of phoneme-size segments. [+phonemic] implies [-indigenous]. This is because the [+phonemic] position is incompatible with implications (2) and (3) of the [+indigenous] position. Certain writers are inconsistent in that they are [+phonemic] and [+indigenous] in the same work.¹

[±generativist]. This refers to whether the writer is committed or claims to be committed to the theory package (See Section 1.4) of generative grammar. [+generativist] implies the following:

- (1) Language universals are pursued and tentative language universals assumed.
- (2) Explanatory adequacy is pursued.

¹ It is interesting to see how they get around the said incompatibility. As an example, I cite Hashimoto 1972, which is [+indigenous] in its Chapter 2, entitled 'Phonetic description', but [+generativist] in its Chapter 3, entitled 'Phonological system'. As we shall see, [+generativist] implied [+phonemic] in those days. Hashimoto does not seem to realize that the chapter on 'phonetic description' is as loaded in phonological theory as the other chapter.

- (3) Binary distinctive features are used and that in formalized ways.
- (4) Synchronic processes are recognized.
- (5) Phonology is taken to incorporate morphophonemics.
- (6) Explicitly laid down descriptive formalism can be applied.

Before the development of nonlinear phonology in the latter part of the seventies, especially before and around the publication of Chomsky and Halle 1968 (hereinafter SPE), [+generativist] should also imply [+phonemic]. (6) opens up the possibility for a writer to apply the descriptive formalism mechanically to the language as a matter of procedure. This is ironical in view of the fact that generative grammar has developed out of a reaction against the post-Bloomfieldian American mechanical "structuralist" idea that adequate description of a language can be achieved by the mechanical application of some "discovery procedure".

[+pedagogic]. This refers to whether the work is written with the teaching of Cantonese to non-native speakers in mind. [+pedagogic] implies the following:

- (1) Phonetic details are attended to.
- (2) The description is subject to the influence of the sound pattern of the native language of the learners.

The first implication is definitely a merit. (2) on the other hand may work both ways. To the extent that it leads to something like contrastive analysis, it is a good thing as it will only deepen our understanding of the sounds of Cantonese. However, if the description is distorted in favour of the sound pattern of the learners' language, it is undesirable.

Apart from these six parameters, which have major implications in the way a linguist regards Cantonese phonology, other variables also bear on the way one handles Cantonese sounds. These include which variety of Cantonese one is describing, how abstract one permits the description to be, and how much phonetic detail one includes.

In this section we have discussed the parameters in general terms. The discussion will help us understand individual works when they are referred to in the next section and beyond. On the other hand, only when individual works are examined in greater detail in the rest of the

thesis will the discussion in this section be fully appreciated.

1.3.2 Survey of literature

Any survey of literature should begin with bibliographic works. In this regard Yang 1981 and Lucas 1985 are extremely helpful. They should be supplemented by Yang 1974, Yǔyán Yǎnjīusuǎo 1978 and 1983, and the yearly issues of Linguistic Bibliography. As I gained access to Lucas 1985 only towards the completion of this thesis, I have not been in a position to make reference in this thesis to the highly relevant Phoon 1976, Tse 1982 and Wong 1982 entered there. I also failed to have access to McCoy 1966, Hashimoto 1971 and Yú 1979. For writings in Japanese I rely on their documentation in Hashimoto 1972.

Phonemic transcription or some kind of sound pattern characterization of Cantonese began in the second half of the 19th century in such dictionaries as Williams 1856 and Eitel 1877. There are a number of reasons for my ignoring these works:

- (1) As Wong (1940:4-5) points out, they are not based on Cantonese proper.
- (2) They are very out-dated.
- (3) Their description of Cantonese sounds is not rigorous.

Among works of last century I only refer to Chan 1899-900 and Ball 1899-900 for their contribution to the understanding of Cantonese tones, to Parker 1880b and Lockhart 1882 for its documentation of neglected syllables, and to Parker 1880a for both reasons.

Seers 1908 suffers from the fact that while his object of description includes other Yuè dialects than Cantonese proper, he does not spell out which dialect a particular statement is directed at. We may safely date the first rigorous description of Cantonese sounds to Jones and Woo 1912. The well-known sinologist Karlgren (1915-26 and 1923) also furnishes a rigorous transcription of Cantonese sounds, in notations borrowed from Swedish dialectology, together with some phonetic description. The following matrix shows the major works since Jones and Woo that describe or include a description of the overall sound system of Cantonese, characterized in terms of the parameters identified in the last section.

	[consv]	[dialt]	[indig]	[phonm]	[gener]	[pedag]
Jones & Woo 1912	-	-	-	+	-	+
Karlgren 1923	+	+	+	-	-	-
Wáng 1936-7	+	+	+	-	-	-
Wong 1940	+	-	+	-	-	-
Chao 1947	-	-	+	-	-	+
Chén & Bái 1958	+	+	+	-	-	-
Yuán et al 1960	+	+	+	-	-	-
Cheng 1968	+	-	+	+	+	-
Kao 1971	+	-	-	+	-	-
Hashimoto 1972	+	-	+	+	+	-
Dow 1972	+	+	+	-	-	+
Ráo et al 1981	-	-	+	-	-	-

Jones & Woo, Karlgren, and Chao¹ are basically original. Wáng, on the other hand, is subject to much influence from Karlgren, though his notation is in IPA. Jones & Woo and to a lesser extent Chao have a good deal to offer in phonetic description, and are unsurpassed today in this respect. Wong, Yuán et al, and Chén & Bái are followers of Jones & Woo, Wáng, and Chao respectively. This is symbolized by the choice of notation for sibilants: *s*, *f* and *ɕ* in that order. Wong and Yuán et al enrich their description with information on variation, casual speech, morphotonemic alternation, orthoepy, etc. As Yuán et al is the first comprehensive description of Chinese dialects and Wong is an authoritative pronouncing dictionary, and since both are written in Chinese (but Wong has an English appendix), both have become standard works among the Chinese today. There has not been any substantial contribution to the overall description of Cantonese sounds since Yuán et al.

¹ The date is omitted for works listed in the table.

Cheng 1968 is the first generativist treatment of Cantonese sounds. The second of the kind is Hashimoto. Hashimoto makes no reference to Cheng, and their treatments are quite different. Both of them are more concerned with the application of the generativist formalism than with descriptive adequacy. In neither work is there much updating for the sake of observational adequacy, much phonetic detail, or much new insight into the sound pattern of Cantonese. Nevertheless, both descriptions have benefited from the superiority of distinctive features. Thus, as Cheng puts it,

most [old] problems are pseudo- problems (sic), in the sense that they pose difficulties only for 'distinctive segment analysis' (....) but not for distinctive-feature analysis.

Apart from being a phonology, Hashimoto is a very useful reference manual. Referring to Hashimoto, Chen (1984) writes:

The bibliography there is nearly exhaustive except for a few unpublished theses. It would be fair to say that Hashimoto (1972) supersedes all previous phonetic and phonological descriptions of Cantonese.

It is just not true that Hashimoto supersedes Jones & Woo or Chao in phonetic description. And Hashimoto's usefulness today does not really lie in its bibliography. Indeed with regard to bibliographic coverage it has been far superseded by the bibliographic works cited above. It is useful more because of its comprehensive documentation. Thanks to it, those who do not read Japanese can have indirect access to Japanese scholarship. Its documentation of scholarship in Chinese is still the most comprehensive today among publications in English. The book contains a plain syllabary, a "morpheme syllabary", and a list of morphemes subject to morphological tone change. Throughout the book there are various kinds of exhaustive or near-exhaustive listing. And apart from synchronic phonology of Cantonese, over one third of the book is devoted to diachronic phonology and to other Yuè dialects. No Cantonese phonologist can afford to miss this reference manual. As a phonology, however, all that the book does is furnish (i) an agglomeration of ideas from previous work, and (2) an exercise in generativist formalism.

Kao's persuasion is "structuralism", mainly of the post-Bloomfieldian mechanical type, but subject also to some influence from the Prague school. Distinctive features are not used, though. She is more eager than Hashimoto to evaluate other descriptions, but she gives too much attention to the phonemic solutions embodied in romanizations used in dictionaries and Cantonese textbooks, which are not meant to be linguistically rigorous. Quantification and graphicization are two salient features about the presentation in the book. Quantification comes from (i) an acoustic study and (ii) a statistical study; both are the first of their kind in Cantonese phonological work.

Dow includes a chapter comparing Cantonese sounds with Mandarin. The service this chapter does to Cantonese sounds lies in its phonetic description. It is the only work that is comparable to Jones & Woo and Chao in this respect, and is complementary to the two. It is a pity that the 1984 revised edition drops this chapter.

Ráo et al's value consists in its relative non-conservatism (as illustrated by its recognition of the rimes $\varepsilon:m/p$ and $\varepsilon:t$ and the wealth of raw casual speech and variation data contained in it. But he has not made the best use of his primary linguistic data in his characterization of the Cantonese sound pattern.

Apart from the above-mentioned works, other works that describe or contain a description of the overall pattern of Cantonese sounds include Cén 1946, Egerod 1956, Wáng 1957, S Cheung 1972, Gão 1980, and Wáng 1985.

Needless to to say, contributions to the description of Cantonese sounds are not confined to these overall accounts. For example, in the area of rime, Hashimoto and Hashimoto 1968 is a trial treatment of Cantonese vowels in generativist terms (superseded by Hashimoto 1972); Light 1977 tries to bridge the gap between indigenous and Western scholarship in the characterization of rimes; and more recently Luke (1983) discusses a small part of the rime system in depth.

The area that has attracted intellectual interest the most is tone

and intonation. Thus, Zōng 1964 and Y Cheung 1969 are efforts to establish the split of tone 1; Fok 1974 and Vance 1977 greatly deepen our understanding of Cantonese tones; and Tse 1978, Gandour 1981 and Ching 1981 provide us with supplementary information on them. Whitaker (1955-6) is a classic account of morphological modification of tone in Cantonese. Fan 1979 explores areas not reached by Whitaker. Yú 1984 presents certain sub-lexical and semi-lexical tonal alternations. Kam 1977 and Wong 1981 tackle diachronic tone change. Killingley (1982, 1985a,b) explores the possibility that Cantonese has five tones instead of six.

The International Phonetic Association 1949 and the acoustic study Lee 1985 are important contributions to Cantonese phonetics. Besides these two works, bits and pieces of information on the phonetic details of Cantonese come from works of diverse nature, including Huang 1965, O'Connor 1980 and Cén 1982, and the acoustic studies Lisker and Abramson 1964, Clumeck et al 1981 and Iwata 1985.

Among the other relevant studies, Zhāng 1983 and Bauer 1984 are efforts to expand (i.e. to update) the somewhat impoverished Cantonese syllabary so far taken for granted. Yeung 1980, Luke 1984 and the other works of Bauer consider socio-phonological variation. Wong 1981 referred to above, and Bauer 1979 and 1983 are written with a view to giving support to the "lexical diffusion" theory of sound change proposed by William Wang. Yip 1980 contains applications of the descriptive device of autosegmental phonology to certain tonal phenomena in Cantonese. Yip 1982 includes an attempt to reinterpret the working of a Cantonese secret language (reported in Chao 1931) in terms of CV skeleton phonology, one branch of autosegmental phonology. Bai 1982 is the only work devoted to casual speech. For Malayan Cantonese one must refer to the works by Killingley.

1.4 Framework of description

I shall make use of binary distinctive features, rules, rule ordering, rule schematism, and autosegmental phonological representation;¹ all this reminds us of (post-SPE) generativist phonology. On the other hand, there is evidence that I have absorbed ideas from Saussure, (Firthian)

prosodic analysis,² and indigenous Chinese phonology. To borrow the words of Wells (1982:xv), my descriptive standpoint "could be said to involve an eclectic amalgam of what seems valuable from both older and newer theoretical approaches."

Admittedly I am in danger of being inconsistent, and a principled defence of eclecticism is appropriate here. In this regard allow me to quote Hudson (1984:126-30) at length:

I prefer to call [a 'theory' in relation to linguistics] a 'theoretical package' because it is just that — a collection of separate theories which are presented as a single package. (A typical package might include theories about phonology, grammar and semantics, plus various other theories about matters such as how we learn languages as children.)

[B]y the 1980s we have an impressively long list of packages (....)

Is it because the theories in each package are so inextricably bound up with one another that you can't accept one without accepting the lot?

[V]ery few assumptions or theories are completely unique to any one package, and the ways in which they are combined often seems fairly random.

[One] reason why linguistic ideas are divided into packages is a social one (....)

[L]inguists tend to present their wares in 'packages' of theories which can be considered on their individual merits to a much greater extent than is sometimes implied (....)

Hudson expresses succinctly how I see the various brandnamed linguistics theories. If my descriptive standpoint involves theory packages, so do the various brandnamed theories in the final analysis. The danger of inconsistency lurks behind eclecticism just as it lurks behind a brandnamed theory. Whether the theory package adopted in a phonological description forms a coherent whole and whether it is in accord with the primary linguistic data are empirical questions and thus cannot be judged a priori. At this point the elaboration of my descriptive

¹ See Goldsmith 1979 and Hulst and Smith 1982, 1984 for autosegmental phonology.

² See Palmer (ed) 1970, Hill 1966 and Chapter 7 of Anderson 1985 for Firthian prosodic analysis.

standpoint might be in order.

The fundamental note of my descriptive standpoint is a break with the principle of unilinear phoneme-size segmentation.¹ This principle has been taken for granted in classical taxonomic phonemics and inherited uncritically by early generativist phonology.

The early development of classical taxonomic phonemics was tied up with the need and desire for the (inevitably unilinear) transcription of utterances. Note that Pike 1947 bears the subtitle *a technique for reducing languages to writing*; and as late as 1957, Jones writes:

[T]he physical view of the phoneme is on the whole better suited to the needs of ordinary teaching of spoken languages and (...) for those who are called upon to reduce to writing languages hitherto unwritten or to improve upon existing unsatisfactory orthographies. (p.192)

In the world of alphabetic writing, transcription of utterances would easily be thought to be synonymous with "alphabetization of sounds". Thus Firth (1948:8) writes:

The development of comparative philology, and especially of phonology, also meant increased attention to transliteration and transcription in roman letters. [T]his might [have] contribute[d] to the tendency, both in historical and descriptive linguistics, to phonetic hypostatization of roman letters, and theories built on such hypostatization.

¹ Note the qualification 'phoneme-size' which makes my characterization not entirely the same as Goldsmith's (1979:17) characterization of the 'standard linguistic assumption regarding the nature of phonological representations' as the *Absolute Slicing Hypothesis*. I am of the view that slicing between syllables is practicable, at least in the case of Cantonese.

Stripped of the straitjacket of the principle of unilinear phoneme-size segmentation and freed from preoccupation with linguistic universals and explanatory adequacy (See Section 1.1), I attach much importance to primary linguistic data, which almost solely determine the framework I adopt. Note that while prosodic analysis shares with me the break from unilinear phoneme-size segmentation, generativist phonology inherited the principle, which has shaped and in my opinion hindered its development.¹ There is little wonder that prosodic analysis and indigenous Chinese phonology come close to my descriptive standpoint, especially in their recognition of the syllable as the primary phonological isolate.

We need a little dialectics to establish that a break with the unilinear phoneme-size segmentation principle is not necessarily in contradiction with the recognition of such phoneme-size segments as onset, vowel and coda: their status is secondary to the syllable and their interpretation depends on their relationship to the syllable.

To facilitate comparison between the theory package for the present thesis and other packages, I introduce a matrix, which compares classical TAXONOMIC phonemics, SPE phonology, PROSODIC analysis, INDIGENOUS Chinese phonology and the framework adopted in this THESIS, in terms of the following set of parameters:

[formal]: whether the theory is form- or substance-oriented.

[process]: whether synchronic processes are recognized.

[feature]: whether distinctive features are used.

[universal]: whether language universals are presumed (and thus explanatory adequacy pursued).

[polysystemic]: whether polysystemicity² is recognized.

¹ Referring to the principle of unilinear phoneme-size segmentation underlying (early) generativist phonology, Hill (1966:223) writes: 'It seems indeed especially unfortunate that a theory (...) in which the parameter rather than the segment is the natural basic isolate, should not have followed out the implications of this for analysis.' He predicts, correctly, that the situation 'is bound to change'.

² Polysystemicity refers to the principle that 'the set of alternances at any specially defined point in the structure is *sui generis*, and need not correspond in formation to the set at another

[syllable]: whether the syllable as a primary isolate is recognized.

[long component]: whether components longer than the phoneme are recognized.

[phonemic]: whether the principle of unilinear phoneme-size segmentation is assumed.

The matrix follows:¹

	TAXONOM	SPE	PROSOD	INDIGEN	THESIS
[formal]	-	+		+	+
[process]		+	-	-	+
[feature]	-	+	-		+
[universal]	-	+	-	-	-
[polysystemic]	-	-	+	+	+
[syllable]		-	+	+	+
[long component]	-	-	+	+	+
[phonemic]	+	+	-	-	-

The matrix shows that what clearly differentiates my standpoint from prosodic analysis and indigenous Chinese phonology is my recognition of synchronic processes and consequently my formulation of rules. This very fact brings my standpoint closer to SPE phonology. However, because of my other differences from SPE phonology, my rules are inevitably different from the SPE type of rules in nature and appearance.

Saussure's ideas have developed into several brands of linguistics structuralism. But I believe that the formal slant, at least at the abstract end, of this thesis is in the spirit of Saussure's principle that

point'(Hill 1966:217). In the context of Cantonese, where phonology is relatively autonomous (and is methodologically assumed as such in this thesis), polysystemicity consists in recognizing different interplay of contrasts between onset and coda.

¹ Certain holes are left unfilled either because the theory holds no committed view on the parameter or because I do not know the theory thoroughly enough to assign the parameter a value.

"the language itself is a form, not a substance." (1982:120, 1922:169) Moreover, it is my view that the idea of distinctive features is a natural consequence of another (related) principle of his, that "in a language there are only differences, no positive terms." (1982:118, 1922:166)

At this point it should be clear that my descriptive framework's partial resemblance to generativist phonology is superficial. One easily overlooks the fact that some of the practices within generativist phonology are not unique to this theory package. Thus, while it is well known that binary distinctive features are the creation of the Prague school, it is not so well remembered that the idea of synchronic processes dates back at least to Bloomfield 1933:213, where the rules are even ordered. The idea of long components, a salient feature of autosegmental phonology (which is held to be generativist), dates back to Harris 1944 and Firth 1948.

Following this last point on autosegmental phonology, I do make use of the descriptive device of the theory. Although autosegmental phonologists usually claim to be or are thought to be generativists, I hold that its descriptive device is open for any linguist who is not imprisoned by the principle of unilinear phoneme-size segmentation. From the [+phonemic] (including SPE) phonologist's point of view, autosegmental phonology constitutes a blow to the [+phonemic] principle and therefore calls for a major revision of theory. From the viewpoint of other phonologists, autosegmental representations are above all an ingenious exploitation of the geometry of phonological representation.¹ Referring to autosegmental phonology, Walton (1983:274) writes:

The theory, however, has surprisingly little to offer, conceptually speaking, to the Chinese case and most likely to the analysis of Sino-Tibetan languages generally. This derives no doubt from Goldsmith's concern with phonological processes rather than surface phonetic description (....)

¹ This is not to deny the far reaching implications of such exploitation.

What Walton calls "phonological processes" in fact include processes that are morphophonological. Indeed I use autosegmental descriptions the most when I am dealing with the rather low-level temporal implementation of segments.

In another context in the same work, Walton writes:

[T]he predominant phonological theories have been and continue to be ill-equipped to characterize the defining features of Chinese and of Sino-Tibetan languages in general, not just because of purely linguistic factors but rather because of the cultural milieu within which these theories have evolved. It seems fair to say that the majority of current phonological theories have been developed within the confines of Indo-European cultural and linguistic tradition, have drawn their impetus from initial work on Indo-European languages, and have then been modified when applied to non-Indo-European sound systems.

He then writes at length to establish what has been summed up by Firth as the "phonetic hypostatization of roman letters". Leaving aside this hypostatization, to follow on this quote of Walton's, it might be thought-provoking to note that autosegmental phonology has developed out of a rethink of the prevalent modes of phonological representation (a kind of tentative formal universal) in the course of analysing African languages in full recognition of their intrinsic characteristics. When more Sino-Tibetan languages are studied in full recognition of their intrinsic characteristics, we expect a further swing away from perennial Eurocentrism in phonology.

CHAPTER 2: THE REFERENCE DESCRIPTION

Descriptions of the Cantonese phonological system abound. No two analysts present the more or less similar raw data in exactly the same manner, adopting the same framework of description with the same set of basic assumptions, either implicit or explicit. Wú 1976 presents a by no means complete table of different notations, employed in twenty-one different works, for the systems of onset, rime and tone. Though Wú speaks only of "notations", different phonemic solutions in fact underlie many of the differences in notation. For example, some systems recognize nine tones while some recognize only six; some (e.g. Chao 1947) align the rime *uŋ* with *ow* while most others align it with *u:*.

With wide discrepancies in the description of Cantonese phonology, it seems an intractable task to present all the major analyses, have sympathetic understanding of each of them, and evaluate them. For one thing, in order to do just that one needs to have a kind of reference description (RD) which is (i) observationally adequate, (ii) concrete enough to be compatible with different kinds of phonological treatment, and (iii) abstract enough to be free from problems of phonetic realizational exactitude.

The present chapter is exactly devoted to developing such a reference description. In Section 2.1, as an initial step, a kind of "standard" description (SD) will be presented as point of departure for developing the RD eventually. SD is the relatively widely accepted core of Cantonese phonological description, shared, subject to minor variations, by several widely circulating works. Then in Section 2.2 it will be shown that revision of SD is necessary in order to satisfy the requirement of observational adequacy. Section 2.3 summarises the result of such revision, which means the presentation of RD itself.

2.1 Presenting a standard description

The SD we are presenting here is a kind of common ground in the description of Cantonese sounds shared by Yuán et al 1960 (and several works under its influence), Hashimoto 1972, Light 1977, and some other works having a [+indigenous] slant. While Hashimoto, following a

procedure not unlike our own in this thesis, treats such a description as the point of departure for further abstraction and discussion, Yuán et al and their followers more or less regard it as a complete phonological analysis in itself and do not go much further. Light's position is intermediate between the two. His article aims at justifying the indigenous analysis of the rime and at the same time tries to reconcile this language-specific orientation with Western modes of description.

The SD that we adopt has the following characteristics:

- (a) The syllable is viewed as a primary phonological isolate.
- (b) The syllable has three immediate constituents, namely tone, onset and rime (bearing in mind that tone, unlike the other two, is suprasegmental).
- (c) The onset is optional.
- (d) The rime has vowel and coda as its immediate constituents.
- (e) The coda is optional.
- (f) Phonetic diphthongs (which are all narrowing diphthongs) are treated as vowel + coda.
- (g) Within the system of coda, -p, -t, -k are regarded as contrastive with -m, -n, -ŋ respectively (Light excepted), implying, if the analysis is to be consistent, that only six tones are recognized.
- (h) The idiosyncratic syllables [ɱ] and [ŋ] lie somewhat outside the system developed so far, and their existence has to be mentioned in passing but cannot be integrated into the systematic characterization of the syllable.

Let S = syllable, O = onset, R = rime, V = vowel and Cd = coda; (a) to (f) can be schematized as the following "formation rules":

- [1] $S \rightarrow T (+ O) + R$ <where T is non-localized>
 $R \rightarrow V (+ Cd)$

(g) and (h) can be implied when the individual terms of the four constituents of the syllable, each a paradigmatic system in itself, are spelt out. Thus:

[2]

T = T1, ¹ IPA ˥S, i.e. high-falling.

T2, IPA ˨˨˥S, i.e. high-rising.

T3, IPA ˨˨˨S, i.e. mid-even.

T4, IPA ˨˨˨˨S, i.e. low-falling.

T5, IPA ˨˨˨˨˥S, i.e. low-rising.

T6, IPA ˨˨˨˨˨S, i.e. low-even.

O = m n ŋ
 b d dz g gw
 p t ts k kw
 f s h
 l
 j w

V = i:/ɪ y:/ u:/ʊ
 ε:/e œ:/ø ɔ:/o
 ɐ
 a:/

Cd = w j
 m n ŋ
 p t k

The foregoing description has to be supplemented by the following statements about phonotactics:

- (a) Only certain combinations of V and Cd give rise to wellformed R.
- (b) Within the R, the distribution of -p -t -k is exactly the same as their homorganic nasal counterparts -m -n -ŋ respectively. It

¹ The motivation for the numbering can be seen in the traditional name given to each tone in indigenous Chinese phonology:

	píng	shǎng	qù
yīn	T1=yīnpíng	T2=yīnshǎng	T3=yīnqù
yáng	T4=yángpíng	T5=yángshǎng	T6=yángqù

'píng → shǎng → qù' is the traditional order, so is 'yīn → yáng'. Granted these orders, the table above could still be read in columns rather than rows, such that the second tone to come up would be yángpíng rather than our yīnshǎng. The order adopted in this thesis is the commoner of the two possibilities. A third way of ordering is also possible: T1, T2, T3, T5, T6, T4, in terms of average pitch height from high to low.

follows that when specifying wellformed R, only one of the two series need be mentioned. That is to say, as far as the phonotactics of the R is concerned, -p need not be distinguished from -m, nor -t and -k from -n and -ŋ respectively.

According to SD, then, the following table exhausts the inventory of R, and gives information as to (i) which "allophone" a V takes in a particular R and (ii) what gaps exist for the V+R combination.

[3]

	-∅	-j	-w	-m/p	-n/t	-ŋ/k
i:	+	-	+	+	+	[ɪ]
y:	+	-	-	-	+	-
u:	+	+	-	-	+	[ʊ]
ɛ:	+	[e]	-	-	-	+
œ:	+	[ø]	-	-	[ø]	+
ɔ:	+	+	[o]	-	+	+
ɐ	-	+	+	+	+	+
a:	+	+	+	+	+	+

[] = V in variant form

+

- = illformed

2.2 Departures from Standard Description

In this section considerations of observational adequacy, as far as present-day Cantonese is concerned, prompt us to identify certain omissions and over-differentiations on the the part of SD, as well as other kinds of undesirability in the form in which it has been presented in the last section.

2.2.1 Missing rimes in Standard Description

As mentioned in Section 1.3.1, the rimes ɛ:w, ɛ:m/p, ɛ:n/t, om/p have repeatedly been omitted from descriptions of Cantonese sounds. This is due largely to the describer's dependency on Wong 1940 as guidance and a prime if not ultimate source with regard to the Cantonese syllabary. Wong 1940 is one of the earliest works published in Chinese to employ the IPA for the representation of Cantonese sounds. As a pronouncing dictionary (rhyming dictionary to be exact), it

RD

supersedes all its predecessors in the coverage of morpho-syllables: 10000 in all (p.2, English section). As such it has been highly influential among later analysts of Cantonese sounds and dictionary compilers. Wong's claim that there are 53 rimes in Cantonese remains largely unchallenged. The omission of the rimes identified above could be attributed to the state of the language at that time — after all what he describes is the Cantonese of half a century ago. Yet two other reasons might also be responsible.

First, Wong 1940 is demonstrably conservative in outlook, reluctant to record innovations in lexical incidence. Except for the most familiar items, the pronunciation provided for each entry is chiefly a projection of the pronunciation given in time-honoured, more or less pan-dialectal pronouncing dictionaries for Middle Chinese,¹ of which the ultimate source is *Qǐè Yùnn*, published in 601 (Cf. Shàn 1980).

Second, to a certain extent all compilers of dictionaries about Cantonese are faced with the difficulty that Cantonese is not a literary language: not every morpho-syllable in Cantonese can be identified with a grapheme, i.e. a Chinese character. Thus, admission of entries in a Cantonese dictionary is very often subject to, or at least misoriented by, the availability of corresponding characters. Hence the easy omission of graphically non-existent or unstable morpho-syllables. In point of fact, all the omitted morpho-syllables resulting from the omission of the rimes in question lack a corresponding graphical representation that is anything more than an "idiolectal" or ad hoc form.

Occurrence of om/p and ε:w has in fact been reported by works before Wong 1940. For example, Parker 1880a and 1880b register the syllables lom, mom, om, bom, pom, gom and kom; Lockart 1882 further registers bop and ɲop; and Jones includes ε:w as one of the "diphthongs". While Wong never mentions om/p, he does refer to ε:w, dismissing it as "not permissible in the best usage of the dialect." (p.5, English Section).

¹ Also called Ancient Chinese. The Chinese of around the 5th century is considered most representative of Middle Chinese.

Fortunately the situation is changing and these missing rimes and their resultant syllables are beginning to be recognized in published works:

- (a) Hashimoto (1972:218) records the rimes $\epsilon:t$ and the syllables $p\epsilon:t^6$ and $b\epsilon:t^6$.¹
- (b) Ráo et al (1981) recognizes the syllables $g\epsilon:m^1$, $g\epsilon:p^1$, $t\epsilon:t^6$ and $k\epsilon:t^1$.
- (c) Zhang (1983) recognizes $g\epsilon:p^6$, $p\epsilon:t^1$, and $b\epsilon:w^6$ in addition to the $p\epsilon:t^6$ of Hashimoto.
- (d) Bauer (1984) recognizes the rime $\epsilon:t$ to be Cantonese proper and the rimes $\epsilon:n$, $\epsilon:p$ and $\epsilon:m$ as "developed under the influence of English"(p.9)

As a native speaker of Hong Kong Cantonese, I attest the grammaticality of the following items:

[4]

SYLLABLE	TONES	GLOSS
$l\epsilon:m$	2	to lick
$g\epsilon:m$	1	(to lose) a game
$k\epsilon:m$	1	camp/chemistry
$k\epsilon:m$	4,2	$k\epsilon:m^4k\epsilon:m^2$ = noise of clearing throat
$d\epsilon:p$	1	to taste appreciatively
$g\epsilon:p$	6	to grip
$f\epsilon:n$	1	friend(ly)
$j\epsilon:n$	1	yen (Japanese currency)
$w\epsilon:n$	1	van, especially mini-bus
$k\epsilon:t$	1	sound of giggling
$l\epsilon:t$	2	$bi:l\epsilon:t^2$ = billiard
$p\epsilon:t$	1	buttocks
$p\epsilon:t$	6	mass (a classifier)
$b\epsilon:t$	6	$l\epsilon:m^4b\epsilon:t^6b\epsilon:t^6$ = soft as mud
$t\epsilon:t$	6	$bi:n^2t\epsilon:t^6t\epsilon:t^6$ = flat
$f\epsilon:t$	6	} $fi:l^4li:l^1f\epsilon:t^6l\epsilon:t^6$ = sound of crying
$l\epsilon:t$	6	
$b\epsilon:w$	6	to jostle with the hip

¹ The rime and the syllables are given in the *Morpheme Syllabary*. It is mysterious that she completely ignores the rime in her own description outside this syllabary.

de:w	6	to throw away
gom	2	in that case...
bom	4	Boom!
lom	4	gow ¹ lom ⁴ lom ⁴ = quite tall
dom	2	sound of something falling into water
bop	4	heart-beat kind of sound
tsop	?	(no intrinsic tone) sound suggesting a swift cut by scissors

I contend that the case for the rimes $\epsilon:m/p$, $\epsilon:n/t$, $\epsilon:w$, and om/p is by now established. This is significant in that they fill some of the gaps that exist in the potential combinations of V and Cd. In light of these newly recognized rimes, the rows headed by $\epsilon:$ and $o:$ in table [3] has to be revised as follows:

[5]	$-\emptyset$	$-j$	$-w$	$-m/p$	$-n/k$	$-ŋ/k$
$\epsilon:$	+	[e]	+	+	+	+
$o:$	+	+	[o]	[o]	+	+

As we said earlier, observational adequacy is a necessary condition of descriptive adequacy. The existence of these rimes thus has a bearing on the nature and details of the rules specifying the wellformed rimes through a filtering system known as "constraints". Both Hashimoto (1972) and Light (1977) attempt to specify such rules. Since neither of them take into account any of the rimes identified above (including $\epsilon:t$ which Hashimoto recognizes elsewhere in her book) when they formulate the rules, their formulations are predictably inadequate. (See Section 4.2.1.1.)

2.2.2 Over-differentiations in Standard Description

Two cases of over-differentiation can be identified in Standard Description, both in the system of onsets. One concerns the opposition $n-$ vs $l-$; the other concerns the opposition between $\eta-$ on the one hand, and $\emptyset-$ (i.e. the lack of onset) on the other. These will be dealt with one after the other.

2.2.2.1 $n-$ vs $l-$

The merger of the onsets $n-$ and $l-$ has long since been noticed to

be under way. According to Chao 1947:18, about one quarter of speakers of Cantonese (presumably in Guǎngzhōu) had lost that distinction. Being the first linguist to report this and other varieties, he should be praised as [-conservative]. However, Barnett (1949:727), a reviewer of Chao 1947, might not agree:

I do not think [Chao] is right in putting so low as one-fourth the number of persons in Canton who have no initial *n*: from long and careful observation I should say that nearly all women and at least one man in four show this feature, and that among men of thirty and below the proportion is much higher.

The fact that a large number of speakers were not able to distinguish between *n*- and *l*- is theoretically more significant than the fact that some speakers did make the distinction. Given that Cantonese is a language shared by a speech community, if *l*- and *n*- are contrastive, no native speaker would internalize a grammar which does not distinguish them at all. The converse is not necessarily true. Cantonese is geographically adjacent to other Yuè dialects and socially interactive with Mandarin. Mandarin and several Yuè dialects do contrast *n*- and *l*-, and that with the same lexical incidence as the theoretical onsets *n*- and *l*- in Cantonese. Moreover a significant number of Cantonese speakers also speak some English, which also contrasts /*n*/ and /*l*/. With all this taken into consideration, it is clear that while the ability to distinguish between [n] and [l] by some speakers may be due to dialectal and/or English influence, the inability to distinguish between the two sounds on the part of some speakers and, what is more indicative, the inability on the part of a presumably even larger number of speakers to identify correctly the (theoretical) lexical incidence of the onsets *n*- and *l*- is inexplicable unless the basic grammar of Cantonese is such that the two sounds are non-contrastive.

Despite the fact that the merger of *n*- and *l*- is branded "wrong" by many prescriptivist writers, the situation in mainstream Cantonese today favours strongly the merger stance.¹ That is to say, at least as far as present-day mainstream Cantonese is concerned, *n*- and *l*- are

¹ Witness the puns (i) *si:2li:w6* 'historical data' vs *si:2ni:w6* 'shit and piss', and (ii) *ma:j5lew2* 'to buy a house' vs *ma:j5new2* 'to buy a button'. I came across a child etymologizing *ləj5di:m3* 'inn' as *nəj5* 'female' + *di:m3* 'shop' and thus speaking of *la:m4di:m3* 'male + shop' by analogy.

basically non-contrastive. While treating n- and l- as separate onsets, Hashimoto (1972:120) makes the following tell-tale observation to the contrary effect:

An interesting phenomenon common among some Cantonese speakers learning English is the confusion of these two initials in the target language, and often English words beginning with [l] are pronounced as with [n], and those with [n] as with [l], which is quite inexplicable, except on the basis of a bias derived from their source language.

Fung 1974 is a dictionary published in Hong Kong with the chief objective of enabling the user to look up the grapheme, i.e. the corresponding character, of a morpho-syllable on the basis of its pronunciation. Presumably because potential users of the dictionary cannot be sure if the onset of a particular item is n- or l-, the compiler adopts the thoughtful measure of juxtaposing every pair of syllables distinguished (theoretically) only by whether the onset is n- or l-, rather than follow the strict alphabetical order which is otherwise how the entries in the dictionary are arranged. This innovative measure in Fung's dictionary is a clear indication of the impracticability of the assumption that Cantonese speakers distinguish n- and l- and know their lexical incidence accordingly.

SD, then, needs adjustment in this regard. n- and l- are non-contrastive. And since [l] is the more likely realization of this merged onset of n/l-, we use "l" in our notation.

2.2.2.2 η- vs Ø-

Another exception to the overall alphabetical order of entries in Fung 1974 is the juxtaposition of η- items with their onsetless counterparts, most likely because of the same reason as for similar treatment for n- and l-.

For historical reasons the distinction between η- and Ø- has extremely low functional load. Two sounds of course do not have to be in complementary or near-complementary distribution in order for them to merge (witness the merger of l- and n-), but as an empirical fact, the merger of η- and Ø-, like that of n- and l-, has long since been noticed to be under way. Chao (1947:21) mentions that except for

RD

interjections, particles, and the proper name prefix a:³, three quarters of the speakers pronounce the onsetless syllables with η - and that one can safely pronounce with η - the onsetless items and the η - items alike. Yuán et al (1960:183) mentions that recently the majority of speakers add η - to the theoretically onsetless syllables and a minority of speakers drop the theoretical η -. This clearly shows that the two sounds are not distinctive for most speakers, for whom the discrepancy lies in the phonetic realization of a single onset, which varies between η - and \emptyset - (which in turn has various ways of actualization).

Pending elaboration when we deal with variation in Chapter 9, we make an adjustment to the SD to the effect that η - is regarded as non-contrastive with the lack of onset. A consequence of this position is that onset is no longer optional: we can simplify the grammar by incorporating the lack of onset into the onset η -.

2.3 Summary

To recapitulate, we have arrived at a RD by making adjustments to SD in accordance with the requirement of descriptive adequacy (with respect to the majority of speakers). The RD can be presented in the following schematic form:

[6]

a. Syllable structure:

$S \rightarrow T + O + R$ <where T is non-localized>

$R \rightarrow V (+ Cd)$

b. Inventory of each paradigm (i.e. system of paradigmatically related entities):

$T = T1$ (hi-fall or HF), $T2$ (hi-rise or HR), $T3$ (mid-even or ME)
 $T4$ (lo-fall or LF), $T5$ (lo-rise or LR), $T6$ (lo-even or LE)

O = m ŋ
 b d dz g gw
 p t ts k kw
 f s h
 l
 j w

V = i:/ɪ y:/ u:/ʊ
 ɛ:/e æ:/ə ɔ:/o
 ʊ
 a:/

Cd = w j
 m n ŋ
 p t k

R:	-θ	-j	-w	-m/p	-n/k	-ŋ/k
i:	+	-	+	+	+	{ɪ}
y:	+	-	-	-	+	-
u:	+	+	-	-	+	{ʊ}
ɛ:	+	{e}	-	-	-	+
æ:	+	{ə}	-	-	{ə}	+
ɔ:	+	+	{o}	{o}	+	+
ʊ	-	+	+	+	+	+
a:	+	+	+	+	+	+

[] = V in variant form

+ = V in basic form

- = illformed

CHAPTER 3: TONE

The present chapter discusses the inventory of tones, questions related to tone modification, and the characterization of tones, dealt with in the following three sections respectively.

3.1 The inventory of tones

Exactly how many tones should be recognized in Cantonese is a matter of some controversy. To this question no definitive answer is available and new proposals are emerging. Four considerations contribute to shaping the answer, namely:

- (1) how syllable-final "occlusion" is handled,
- (2) whether T1 is considered to have split into two tones,
- (3) whether T3 and T5 are considered to have merged, and
- (4) how tone modifications are handled.

Considerations (1) to (3) will be discussed in this section one after another. Consideration (4) is so complicated and its implications so far-reaching that it deserves in-depth coverage in its own right. It will thus form the subject matter of Section 3.2.

3.1.1 Syllable-final occlusion

By grouping together syllables checked by a stop (i.e. -p, -t or -k) under the name of "occluded syllables", a widely accepted pattern of distribution of the various tone shapes can be represented in the form of the following table:

[1] TONAL CODE	TONE SHAPE		ADDITIONAL TONAL CODE
	Plain syllables	Occluded syllables	
T1	high-fall	high-even	T1'
T2	high-rise		
T3	mid-even	mid-even	T3'
T4	low-fall		
T5	low-rise		
T6	low-even	low-even	T6'

If the three tone shapes in the context of occluded syllables are recognized as tones in their own right, then nine tones obtain, i.e. six "plain" tones plus three "occlusive" tones.

The situation can be better appreciated by looking at the correspondence between certain Middle Chinese (MC) categories and present-day Cantonese (PCant) categories. MC recognized a four-way phonological contrast distinguished by laryngeal effects, including pitch as a function of time. The contrast, known as *shēng* in Chinese, is translated as "tone" in English and is thought of as such. The four MC *shēngs* and their PCant reflexes are as follows:

[2]	píng	shǎng	qù	rù
TRANSLATION:	even	ascend	depart	enter ¹
SUBSTITUTE LABEL:	I	II	III	IV
PCan REFLEXES:	T1, T4	T2, T5	T3, T6	T1', T2', T3'

If T1', T3' and T6' are recognized as tonal categories in their own right, forming a class of "occlusive tones", the MC *shēng* IV will then be transparent. But MC IV is in fact recoverable whether or not we posit the three occlusive tones: it corresponds to PCant occlusion, i.e. whenever -p, -t or -k is present. Thus it seems fair to say that the temptation in adopting the occlusive-tone position is not just the transparency of MC IV but the correspondence between a term in MC system of *shēng*/tone (IV) and terms in the PCant system of *shēng*/tone (T1', T3', T6'), which correspondence means great convenience in diachronic studies and inter-dialectal comparison.

If that were the only reason for the establishment of occlusive tones, then it would be easily dismissed in a phonology of PCant studied in its own right, because such a phonology is not committed to reflecting diachronic correspondence. Recall that within the rime the distribution of -p, -t, -k is exactly the same as their homorganic nasal counterparts -m, -n, -ŋ respectively. In other words plain and occluded

¹ The translation 'enter' explains why what I shall call 'occlusive tones' in this thesis are called 'entering tones' by some writers.

syllables echo each other. Occlusion of a syllable, i.e. the switching of -m, -n, -ŋ to -p, -t, -k is then the segmental correlate of occlusive tones. The segmental opposition -m -n -ŋ vs -p -t -k and the tonal opposition T1 T3 T6 vs T1' T3' T6' imply each other. It follows that there is no need for both oppositions to be treated as equally basic. Thus, the thoroughly [+phonemic] phonologist recognizes a primary distinction between /m n ŋ/ and /p t k/ and treats T1' T3' T6' as allotones of T1 T3 T6 respectively, while the thoroughly [+indigenous] phonologist recognizes a primary distinction between T1 T3 T6 and T1' T3' T6' and treats -p -t -k as "co-allophones" of -m -n -ŋ (in the system of coda).

The two analyses seem symmetrical and the difference seems to follow from a difference in point of view. A consideration of the motivatedness of the "processes" involved, however, helps resolve the indeterminacy. -p -t -k on the one side and -m -n -ŋ on the other fall into natural classes in that the former are voiceless stops and the latter are (voiced) nasals. The tonal characteristic that groups the occlusive tones together is their shortness. In terms of direction of determination, the occlusive-tone oriented analysis comes down to [3] and the alternative analysis to [4]:

[3] Short tone → voiceless stop coda

[4] Voiceless stop coda → short tone

Whereas [3] is arbitrary, [4] is phonetically motivated in the sense that pitch can only actualize as long as the segment is voiced.

Another consideration that also disfavours the occlusive-tone oriented analysis is that besides T1', T3' and T6', arguably T2' and marginally T4' also exist (see Section 7.2.1.1). Faced with this situation, the occlusive-tone oriented analyst is obliged to expand the inventory of tones from nine to ten or eleven. On the other hand, all that other phonologists have to do is to relax the combination restriction between stop codas and tone. If inventories of contrastive entities are held to be more fundamental than combination restrictions, then the emergence of T2' and T4' renders the occlusive-tone oriented analysis even less appealing.

If the occlusive-tone analysis is not as adequate as the occlusive-

coda analysis, it is more consistent and less redundant than the occlusive-tone-cum-occlusive-coda analysis, i.e. one that treats -m -n -ŋ vs -p -t -k as contrastive and recognizes occlusive tones at the same time (e.g. Chén and Bái 1958). The undesirability of this last kind of analysis lies not only in its inconsistency and redundancy but also in the additional need to account for the non-occurrence of non-occlusive codas with occlusive tones and of -p -t -k with plain tones.

It should be noted that the direction of determination (or phonetic motivation) argument holds even if the occlusive tones are labelled "checked" tones as in Light 1977:88, where "checked" is presumably used in Jakobson's way denoting glottalization (but in a broadened sense). If "checked" means no more than shortness, then the direction of determination argument applies in the same way as before. If, on the other hand, it is taken to refer to some kind of command effecting a switch from -m -n -ŋ to -p -t -k, then it is no longer a property of tone, which, unlike shēng, has to do with the pitch-time graph only. One might want to argue that the whole thing reduces to the definition of tone. To a certain extent it does.

Light's use of "checked" represents a reluctance to follow the thoroughly [+phonemic] analysis of treating -m -n -ŋ as contrastive with -p -t -k. But his treating "checked" as a property of tone is not compatible with the accepted conception of tone in linguistics literature. Sharing Light's reluctance, I propose to exploit my [-phonemic] position and treat occlusion as neither a property of tone nor a property of coda, but a property of the syllable as a whole. Thus, in addition to tone, onset and rime, which are commonly recognized as the immediate constituents of the Cantonese syllable, a two-term constituent [*occlusion] is recognized, which is extracted either from the nine (or more) tones of the occlusive-tone analysis or from the eight codas of the occlusive-coda analysis. A comparison of the four different positions covered in this section follows:

[5] POSITION	NO. OF TONES	[*occl] POSITED?	NO. OF CODAS
(a) Occlusive-tone	9+	No	5
(b) Occlusive-coda	6	No	8
(c) (a)-cum-(b)	9+	No	8
(d) This thesis	6	Yes	5

The position adopted in this thesis for the treatment of occlusion, which can be labelled as "prosodic occlusion", is at least as adequate as the occlusive-coda position. It avoids such problems as direction of determination, definition of tone, and unstable inventory involved in the occlusive-tone analysis. In addition it has the pragmatic advantage of bridging the gap between the occlusive-tone and occlusive-coda analyses, and by correlation also between the [+indigenous] and [+phonemic] positions, enabling easy comparison with all existing analyses.

3.1.2 Suggested merger of T3 and T5

Merger of T3 and T5 has been suggested, and that by one single writer, namely Killingley. It should be noted that the variety of Cantonese she speaks is Malayan Cantonese, and it follows that what she describes about Cantonese applies to that variety of Cantonese. Yet she believes that the discrepancy between her five-tone position and all others' position (with six or more tones in the case of occlusive-coda analysis) does not arise from the inherent difference between Malayan and mainstream Cantonese:

An immediately obvious explanation [for the discrepancy] would be that both Mainland and Hong Kong Cantonese have one tone more than Malayan Cantonese. But this would ignore the fact that in earlier descriptions of Malayan Cantonese (e.g., Chiang c. 1940), at least nine tones have been proposed. These nine tones are like the antecedents of the present-day six tones attested for Hong Kong Cantonese. The difference between my five tones and other modern writers' six tones seems to lie not in the difference in our accents but merely in the difference between our description. (Killingley 1983:3)

I do not see the "earlier"¹ nine-tone system for Malayan Cantonese has anything to do with the present issue. All we can say is that mainstream Cantonese and Malayan Cantonese used to have the same tonal system, but nothing prevents the tonal system from developing in

¹ Killingley seems to assume that the 9-tone analysis reflects an older version of Cantonese. But as we have seen in the last section,

20th century Cantonese lends itself to both 6-tone and 9-tone analyses. In particular, the 6-tone analysis by Jones and Woo (1912) predates certain 9-tone descriptions.

different directions in the two varieties of Cantonese. She raises all kinds of far-fetched reasons to justify her conjecture. Thus:

I have been never able to get Mainland and Hong Kong speakers to produce orally six minimal tone distinctions to prove the phonemic status of six tones, using free word forms alone. Impressionistically too, (...) [these speakers'] tone and intonation systems do not sound at all foreign (...) (1983:3)

If my T3 and T5 are both free variants of a single Malayan Cantonese tone, then it is only natural that mainstream Cantonese tones do not sound foreign to her. And the fact that she is not able to get speakers to produce minimal sextets does not indicate anything at all, for it involves so many factors. First, the usual way of setting up contrasts is by means of a chain of related groups of words exhibiting minimal contrast rather than by presenting an n-way minimal contrast once and for all (where n is the total number of contrastive elements in a system). So a six-way minimal contrast would be a wonderful bonus for a six-tone system, but is too extravagant a requirement. And even minimal pairs are sufficient but by no means necessary conditions for establishing a contrast. Second, even if a six-way minimal contrast of tones exists, as it does in mainstream Cantonese, its production offhand is a feat for phonologists, not laymen. Third, her requirements for "free word forms" are over-harsh. For example, she does not permit any reference to or indeed any association in the mind with written characters. But reference to characters is a recurrent topic of everyday speech for mainstream Cantonese. This is inevitable where Chinese is the main written language, unlike in Malaysia where Cantonese speakers do not usually read or write Chinese. To the extent that a monosyllable can be used as an answer to the question "What character is this?", we have no reason to dismiss it as a bound form. On top of this, the suffix dzi:⁶ "character" can be attached to any monosyllabic item with a corresponding character, forming a bisyllabic word which is undeniably a free form and constitutes an even more complete answer to the question just cited.

Having said all that, it may still be the case that six-way minimal tonal contrasts of monosyllabic free word forms in accordance with Killingley's requirements do exist. The following is one example:

[6]	jəw	T1	"to worry"
		T2	"oil for machines"
		T3	"fine"
		T4	"to swim/oil for cooking"
		T5	"to have"
		T6	"again"

And this is precisely the syllable Vance (1977) uses in his experiment on tonal distinctions in Cantonese.

Killingley (1983:3) suggests that other writers "are over-differentiating between two allotones and are treating them as two tones (their low-rising and middle level tones) where [she] treat[s] them as one". This shows that she understands very well that what is really at issue is whether T3 and T5 have merged, not the general question of whether there are five or six tones. It is therefore strange that she does not concentrate on establishing the merger of these two tones, but rather tries to show that "where there is no ambiguity of meaning, in theory, [T5] can take on the phonetic pitches and contours of any of the [other] tones" (1985b:12). She goes so far as to make the following comment:

[T]he tone space of any given tone expands or contracts according to the number of register distinctions which are phonologically significant for any given syllable. If there is only one permitted free form with a certain syllable structure (...), it has freedom to move through the entire pitch range (...) without fear of 'bumping' into any other lexical form with the same syllable structure. (1985:12)

Along this line of reasoning, just because /u:/ is the only vowel that occurs in the environment /tj__n/ in English,¹ it would follow that /u:/ can freely exploit the entire vowel space! For this and other reasons, the experiment in her 1985a is only marginally relevant. In order to establish the merger of T3 and T5 in Hong Kong Cantonese, she should have asked a Hongkonger to pronounce the test items, and have differentiated T3 and T5 whenever the distinction applies. Instead, she pronounced the items herself. It is here that the idea of tone space is most relevant: if Hong Kong Cantonese has one tone more than Malayan Cantonese, her Malayan Cantonese tones are predictably not realized exactly as any five of the Hong Kong tones. Not only does the

¹ I owe this example to J. C. Wells.

experiment fail to focus on whether Hong Kong T3 and T5 have merged, it is not exactly an experiment on Hong Kong tones. It can more aptly be described as a study on Hongkongers' reaction to Malayan Cantonese tones. One case is particularly revealing. The syllable *jew* illustrated in [6] above is actually one of her test items. Instead of making a sextet out of this syllable and getting them pronounced by a Hongkonger, she made a quintet (conflating T3 and T5, thus resulting in a polysemous form) and pronounced them herself.

The first sentences in Killingley 1985b:1 read:

The theory that Cantonese has six phonological tones, held by linguists today, can be ultimately traced back to (...) Jones in 1913. However, the non-Cantonese specialist usually assumes that this theory is independently held by Chinese linguists and that it has been tested by up-to-date methods.

Jones and Woo are hardly particularly responsible for the view that T3 and T5 are contrastive. Over this point they share the view of many analysts both before and after them. It is difficult to believe that the contrastiveness between T3 and T5 is not independently subscribed to by a large number of linguists who are native speakers of Cantonese. Killingley's bibliography does not include a single work written in Chinese. This might account for her speculation.¹ Nor do I share the doubt that Cantonese tones have been tested by up-to-date methods; witness Fok 1974 and Vance 1977, which are perception experiments, not acoustic descriptions that, as Killingley observes, "can only serve to measure the physical properties of tone".

Though there exists lexically highly selective variation between T3 and T5 in Hong Kong Cantonese at least (see Section 9.4.3), the overwhelming contrastiveness between the two is beyond doubt in present-day mainstream Cantonese.

3.1.3 Suggested split of T1

According to our reference description, T1 is a high-falling (HF) tone. This is a simplified account, for a high-even (HE) variant has long

¹ Note that even Jones worked in collaboration with a native speaker of Cantonese.

since been noticed. To treat HE as a variant of T1 is to say that HF and HE are mutually non-contrastive. But this is not a position held by all Cantonese phonologists. Thus, Zōng 1964 and Y Cheung 1969 argue along the same line that HF and HE have each become distinctive tones, effecting a split of T1. This position is also shared by Yú (1979, 1984). As a result of the split, the total number of tones would become seven. In the case of the occlusive-tone oriented analysis, since HF is not available for T1' the total is again increased by one only.

The tone-split position might have something to do with the writers' phonological standpoint. For instance, implicit in Zōng's and Y Cheung's argument is the principle that non-contrastiveness implies either free or conditioned variation. Since at least in certain cases the distribution of HF and HE is neither one of free variation nor one of conditioned variation, HF and HE must be, according to Zōng and Y Cheung, deemed to be contrastive. By "conditions" they mean strictly phonetic ones. While I share with them the insistence on morphology-free phonological contrasts, it seems that Zōng and to a lesser extent Y Cheung have not given much consideration to the built-in variability in the sound pattern. Thus, besides the free and (strictly phonetically) conditioned variations they recognize, there are also regional, chronological, inter-personal and stylistic variations; and between the optionality of free variation and the obligatoriness of strictly conditioned variation, there are variable processes which depend on non-linguistic considerations.

The long list of minimal pairs furnished by Zōng and Y Cheung embodying the HF vs HE distinction, however, cannot be disregarded. As a native speaker of Cantonese myself I do not share their judgment of contrastiveness between those pairs; nor do Ráo et al (1981:276-8). But the pairs are sufficient to establish the contrastiveness of HF vs HE in their idiolects. On the other hand, the fact that the majority of writers, before and after Zōng 1964, maintain T1 to be an unsplit tone has prompted me to treat the one-tone position as basic and the split-tone position as idiosyncratic and transitory.

T1 split is a complicated issue. Apart from the community/idiolect factor, it is also intermingled with the tone sandhi involving HF and HE in both Zōng's and Y Cheung's account. For Y Cheung, there is the

added complication of stylistic variation. For these reasons, the issue will be picked up again in Section 3.2.2, when the transitory nature of the tone-split can be viewed in a wider perspective, and in Section 9.2.5 in the context of the discussion on variation.

3.2 Tone modification and related questions

Tone modification is not a unitary phenomenon. It is rather the result of the action and interaction of a number of different forces in operation, namely allophonic variation, tone sandhi, tonal morphology, lexically selective tone change, elision, "tone stability",¹ tone neutralization, and tone "coercion".² Work in the past has suffered from the failure to identify each of these different forces in action, thus giving rise to much confusion. For the sake of clarity of exposition I divide this discussion of tone modification and related phenomena into parts, dealing respectively with (i) the high-rising modified tone (HR*) and T2, and (ii) the high-even modified tone (HE*) and T1.

3.2.1 The high-rising modified tone and T2

A morpheme is sometimes realized not in its "basic" tone but in a tone shape that is clearly high-rising. In this section we are concerned with the alternation between HR* and (some of) the basic tones. Three different characterizations of HR* have been put forward, depending on (a) whether HR* is considered identical with the basic tone T2, i.e.

[*identical], and

(b) whether HR* is considered variable or invariant, i.e. [*variable].

The three views of HR* can be characterized in terms of these two variables as:

- (i) [+identical] (implying [-variable])
- (ii) [-identical, -variable]
- (iii) [+variable] (implying [-identical])

The [+identical] view is simple and straightforward. Though they do not make any reference to HR*, Jones and Woo (1912) in essence take this view, since all expected occurrences of HR* are given there simply as T2. Identifying this high-rising modified tone with T2, they are free

² 'Coercion' will be explained later in this chapter.

¹ 'Tone stability' denotes the process whereby some tone-carrying segments^{delete} while the tone itself remains.

³ The asterisk '*' after a tone code (e.g. T1*) or tone shape (e.g. HE*) represents modification.

to refrain from mentioning the existence of the tonal alternation in question, which is quite justifiable for a phonetic reader. The crucial question is whether HR* is really identical with T2. Express mention of this identity is made by Y Cheung (1969:96) and Hashimoto (1972:93-7). The former also makes particular reference to the disparity between Chao's (1947:34) [-identical] characterization (based on Guǎngzhōu Cantonese presumably) and the facts of Hong Kong Cantonese.

The [-identical, -variable] view, as far as I know, originates in Wong 1940. Wong describes HR* as "similar to [T2], yet apparently uttered with a little greater strength"¹(p.362). Chao (1947:34) further spells out the tone shape of HR* as "25" in his own five-level system of tone transcription² as distinguished from the "35" of T2. Yuán et al (1960:189) characterize HR* as having a pitch a little higher than T2.

The [+variable] view is most interesting, and its elucidation sheds light on the proper status of HE* (to be discussed in the next section). This view originates from Parker (1880a:366), who holds that "[b]esides the nine regular Cantonese tones, there are, in short, nine corresponding variable tones". Fundamentally subscribing to Parker's [+variable] view, Ball (1899-900) does not carry the view to that extreme. He finds some of these nine derived tones collapsible by way of neutralization. The first-order neutralization involves the derived occlusive tones with the exception of T1'* (the derivative of T1'). Thus T3* and T3'* are deemed identical, so are T6* and T6'*. The exclusion of T1'* is attributed to the "crescendo effect" present in T1'* but not in T1*, which might in turn be due to the falling contour underlying T1 but not T1'. He posits a second stage of neutralization, which involves T2 (HR) and T3 (ME) on the one hand, both regarded as starting from middle pitch, and T5 (LR) and T6 (LE) on the other, both regarded as starting from the same lowish pitch. Ball also hints at the possibility of still further neutralization, i.e. the approximating of all other rising modified tones to T4*, which has the steepest gradient of all:

[T4*] is so marked and distinctive in its character that it has hitherto well-nigh monopolized the attention and taken the other

¹ My translation.

² See Section 3.3 below for an explanation of this system.

variant rising tones under its own name, or at all events the distinction between these five or more rising variant tones has not been pointed out or clearly defined and they have all been considered by many as one and the same tone. (p.221)

Drawing heavily on Ball, Whitaker (1955-6) recognizes four variant forms of her "2nd modified tone", which corresponds to our HR*. The four variant forms result from Ball's second-order neutralization, and correspond respectively to T1', T2/3, T5/6 and T4, while the plain T1, which does not rise, constitutes HE* rather than HR*. Whitaker also subscribes to Ball's suggestion of the possibility of approximating other rising modified tones to T4, deriving a unified HR*. Figure [7] depicts the series of tone neutralizations as suggested by Ball and Whitaker.¹ The series of neutralizations ultimately leads to the levelling of all tones so derived except T1 (but not T1'), which, according to Ball and Whitaker, does not have a rising tone shape.

[7]

Parker	Ball	Ball & Whitaker	Ball & Whitaker
Pre-neutralization	1st-order neutralization	2nd-order neutralization	3rd-order neutralization
T1*	T1*	T1*	T1*
T1'*	T1'*	T1'*	Unified HR*
T2*	T2* }	T2/3*	
T3* }	T3* }		
T3'*		T5/6*	
T5*	T5* }		
T6* }	T6* }		
T6'*		T4*	
T4*	T4*		

Now this clearly represents a bridge between the [-identical, -variable] view and the [-identical, +variable] view. Accepting the synchronically dynamic nature of tone modification as depicted above, the two views are not necessarily incompatible: while the [-identical, -variable] view recognizes only the maximally neutralized HR*, the [+variable] view is apparently a fuller account of the entire process, with the various possibilities provided for.

¹ The tones are arranged here in the order of average pitch height. This measure helps to highlight the pitch of one tone relative to others, and will be adopted from time to time in this chapter.

Happy as the situation looks, the questions remain to be asked whether the [+identical] view also has some justification and how accurately figure [7] captures the hierarchy of neutralization. To answer these questions, we have to look, albeit briefly, at the function of tone modification.¹ The crucial thing to note is that there are two essentially different kinds of tone modification involved in HR*, which must be identified and kept apart before we can acquire a clear perspective of HR*, and indeed of HE* and T1.

The first type of modification concerns the switching of a lower tone than T2, i.e. T3/4/5/6, to T2. I label it "T2 Switch". The switch signifies a number of things in different contexts. We can, therefore, identify different processes at work involving basically the same kind of tonal switch. Some of these processes are more productive than others. Of the more productive ones, that which involves the adjectival construction Adj₁+Adj₁+dej² "fairly Adj₁" is often cited. Thus, in the environment /__dej² a monosyllabic adjective has its segmental part copied while the second syllable of this reduplicated adjective acquires T2. Alternatively, we can say that the second syllable of the reduplicated adjective switches from a lower tone to T2. The most oft-cited T2 Switch, however, involves a virtually non-productive process, whereby the final syllable of a nominal free form has its tone switched to T2, signifying what Chao (1974:34) summarizes as "that familiar thing (or person, less frequently action) one often speaks of."² For example:

[8] pow ³	"shop"	nej ⁵ pow ²	"rice-shop"
dzæ:k ³	"bird"	dzæ:k ²	"bird" (free form)
fɔ:ŋ ⁴	"room"	fɔ:ŋ ²	"room" (free form)
jeu ⁵	"friend"	sy:ljeu ²	"school-mate"
dæw ⁶	"bean"	dæw ²	"bean" (free form)

¹ The kind of modification which gives rise to HR* is lexically selective and morphologically conditioned. In this phonological study we need not go into the functions of such modification except in order to clarify the status of the resultant HR*.

² A special case of this process is the productive switch of a monosyllabic surname of T4/T6 to T2 in the environment low⁵/[a:]³__, resulting in forms of address loaded with the signification of familiarity, e.g.:

low⁵/[a:]³ { tsun⁴→²
dzi:w⁶→²
mek⁶→²

gok6 "bureau" sy:1gok2 "bookstore"

Since the "familiarity" T2 Switch is only weakly productive, and since the derivation is often coupled with additional meaning apart from "familiarity", two things follow. First, it is not easy to demarcate items related by this process and items (like $\eta\alpha:4$ "tooth" vs $\eta\alpha:2$ "serration") exhibiting similar tonal relation that are related only historically and graphically (i.e. sharing a character)¹. Second, the T2 version cannot help but begin to lexicalize. These two things are in fact related. Given the "familiarity" status of the T2 version, there is a fair chance for it to be used more frequently, and be come across by children for the first time earlier, than the unswitched, T3-6 version. The T2 version thus has good reasons to be lexicalized rather than derived ad hoc. An example is $t\alpha:\eta^4$ "sugar" vs $t\alpha:\eta^2$ "sweet (candy)". In the extreme cases, which are by no means rare, the unswitched version falls out of use, when lexicalization of the T2 version applies to the entire speech community. $t\alpha\alpha:\eta^{4\rightarrow2}$ "orange" and $j\alpha\eta^{4\rightarrow2}$ "fine hair, down" are two items that have just reached the stage of complete take-over by the T2 version. The pair $\eta\alpha:4/2$ cited above represent lexicalization of the T2 version for yet another reason: the specialization in meaning of the T2 version is so great that the connexion between the two forms is no longer transparent.

For these results to obtain, however, the derived tone must be truly indistinguishable from T2. Since lexicalization of the derived T2 does occur from time to time, and since the demarcation between pairs of items related by "familiarity" modification and those that are related only historico-graphically is indeed fuzzy, there is a good case for treating this derived tone as identical with T2. Such treatment is also in accord with the intuition of native speakers of Cantonese, including myself. The following widely recognized puns serve to illustrate the point:

- [9] dzaw¹sen¹ { si:6→2 "Justice of Peace dzaw¹"
 si:2 "with shit all over the body"
- { si:4→2 "in the past"
 si:2 "piece of shit"¹ (P.T.O.)

¹ For example Kam's (1977:195-7) list of alternations representing the 'derivation of [T2] morphemes with specialized restricted meaning' includes both kinds of relation.

The same kind of intuition or observation is also reflected in works geared to the teaching of Cantonese pronunciation. In addition to Jones and Woo (1912) mentioned above, Lau (1972:xxv) also equates the realization of this modified tone with T2.²

If it is now clear that the real-life HR* we have seen so far are identical with T2, it would be too hasty, though, to conclude that the [+identical] view is correct and the [-identical] views incorrect, for there is another kind of HR* which presents an entirely different picture.

Recall that, for the [-identical, +variable] view, the modification of tone consists in having the end-point of any given input tone raised to the top of the pitch range. The fact that the tone after modification can somehow retain the identity of the input tone points to the possibility or even desirability of an Item and Arrangement (IA) analysis (Hockett 1954). In an IA analysis, the modification involved (which bears on HR* and HE* alike) is viewed as the addition of a high tail to the input tone. That is to say, on a single syllable there lies a tone plus something tonal, probably another tone. At least we have to say that the syllable has a complex tone. Let S be the segmental part of a syllable and H be some high pitch; the repertoire of complex-toned syllable will be ST1+H, ST1'+H,...ST6'+H.

Whitaker (1956-7) gets close to such a formulation. She capitalizes on Simon's (1953:xx) speculation on the etymology of HE*, namely a

¹ Y Cheung (1969:96-7) lists seven pairs of homophones between HR* and intrinsic T2 items in the course of establishing the non-distinction between HR* and T2. While the mere claim of their homophonous status begs the question of whether HR* and T2 are identical, the list includes three oft-cited puns, which are more forceful. The puns are gew⁶si:² (already referred to above), dzɔ:⁶dow² and ma:j⁵lɛw².

² Ráo et al (1981:280) are typical of a number of writers/compilers who waver between locally (i.e. on Chinese mainland) authoritative descriptions and their own experience. Thus while in the explanatory appendix they simply paraphrase what Yuán et al have said previously in regard to HR*, 'having a pitch a little higher than T2' (my translation), they nevertheless transcribe the HR* as a switch from the lower tones to T2 in the dictionary entries.

tonal modification arising in lieu of an original suffix $ji:4$ (cognate of Mandarin 6r) "son", comparable and parallel to the Mandarin syllable-final retroflexion. This in turn rests on another speculation whereby $ji:4$, despite its extremely low pitch nowadays, is assumed to have been pronounced as a high even tone in the past. Whitaker believes that this "gives also a plausible explanation for" HR* (p.195). Leaving aside the plausibility of this speculation, an etymological account as such has no bearing on the synchronic description of HR*. We have ruled out on grounds of real-life language use the non-identity between the output tone of certain processes and T2. Hence the name T2 Switch. Nevertheless, irrelevant as the etymological speculation is to T2 Switch, it does seem to have inspired Whitaker to go on the right track in the synchronic description of the complement of T2 Switch within the entire universe of HR*, that which involves complex tones. When Whitaker (1956:203) comments:

Apart from the extremely frequent elision of the suffix $[ji:4]$ modification also occurs in lieu of the utterance of certain words which the speaker chooses to omit, such as $[dzo:2, hi:w^1, jet^1]$, etc. The modification of tones in such cases may then be said to compensate for the omission of these words.

ironically she is trying to extend the uncertain, speculative etymology of familiarity T2 Switch to the much more certain and productive synchronic process of elision of the segmental part of a syllable. The elision in question applies to a handful of items bearing either T1 (incorporating T1') and T2, the only tones that reach the top of pitch range, e.g. $dzo:2$ (perfective aspect marker),¹ hej^2 (locative preposition "at"), and jet^1 (literally "one", signifying various things in various constructions). The synchronic reality of the process under discussion is seen in the fact that elision is optional, such that the unelided form and the corresponding elided form exist side by side. Thus, in relation to the elision of $dzo:2$, people do say $tsøj^1 dzo:2$, etc. as well as the following:

[10]	SYLLABLE	TONK	GLOSS
	$tsøj$	1+2	to have blown
	sek	1'+2	to have known

¹ Whitaker (1956:203) mentions $hi:w^1$, an item not used today, besides $dzo:2$, which means the same.

sej	2+2	to have died
høj	3+2	to have gone
ja:k	3'+2	to have eaten
løj	4+2	to have come
ma:j	5+2	to have bought
ma:j	6+2	to have sold
sek	6'+2	to have eaten

In relation to the elision of høj², People do say dzøj¹ høj² sy:3, etc. as well as the following:

[11]	dzøj ¹⁺²	} sy:3	"to put here"
	dzek ^{1'+2}		"to be accumulating (here)"
	dæn ²⁺²		"to put here"
	fæn ³⁺²		"to be lying (here)"
	tʂa:p ^{3'+2}		"to insert (something) here"
	wæn ⁴⁺²		"to faint (here)"
	tʂɔ: ⁵⁺²		"to be seated (here)"
	dop ⁶⁺²		"to be standing (here)"
	ŋɔ:k ^{6'+2}		"to be (here) with the head raised"

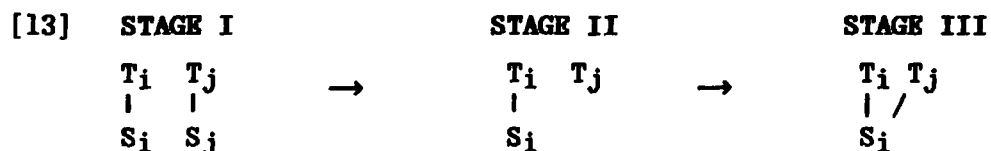
In relation to the elision of jət¹, people do say tsøj¹jət¹tsøj¹, etc. as well as the following:

[12]	tsøj ^{1+HE} tsøj ¹	"to blow a little"
	gət ^{1'+HE} gət ^{1'}	"to pierce a little"
	ga:w ^{2+HE} ga:w ²	"to stir a little"
	gi:w ^{3+HE} gi:w ³	"to call a little"
	dzi:p ^{3'+HE} dzi:p ^{3'}	"to fold a little"
	læ:n ^{4+HE} læ:n ⁴	"to measure a little"
	ly:n ^{5+HE} ly:n ⁵	"to warm to a little"
	mæn ^{6+HE} mæn ⁶	"to ask a little"
	dok ^{6'+HE} dok ^{6'}	"to read a little" ¹

¹ The jət¹ used here is just one of its various uses that are susceptible to the same process of syllable segments elision. Other uses are in the following environments:

- (a) jət¹ Classifier_i — Classifier_i 'one by one'
- (b) Adjective_i — Adjective_i 'very Adjective_i' (where Adjective must be monosyllabic)
- (c) jəw⁵ — 'there is a ...'/'to have a...'

It is in these forms in [10] [11] and [12] where both the elided syllable and the syllable preceding it have the tone maximally retained that we see most clearly the phenomenon of "tone stability" with the resultant realization of two successive tones on one syllable. The descriptive device of autosegmental phonology can best represent and explain the processes involved. The entire modification reveals itself in a three-stage representation:



Stage I represents the initial configuration of the tones and (the segmental part of) the syllables. Stage II sees the deletion of the second syllable. At stage III, the desegmented tone, now floating, gets associated with the preceding syllable, either by universal principle or as a language-specific rule.

In the resultant form, where two successive tones get realized on a single syllable, not only is the tone of the preceding syllable retained but the desegmented tone (T1 or T2) can also be identified provided the speaker so wishes. The exact identity of the desegmented tone, whether it is T1, T1' or T2, has little functional value, and it is subject to neutralization as simply a high tail (H), which is the much more common realization of the second element of the complex tone.

Unlike the identity of the second element, the identity of the first element of the complex tone is sometimes crucial, as in the pair $ma:j^{5+2}$ "to have bought" and $ma:j^{6+2}$ "to have sold". However, this is not to say that neutralization of the first element of a complex tone never happens. On the contrary, smoothing or corner cutting of the contour of a complex tone is a common phenomenon, suspending some of the possible contrasts in the first element of the complex tone, and thus reducing the number of contrastive complex tones. Neutralization of complex tones is a phenomenon of connected speech, especially casual speech. In fact even in cases where the complex tone is maximally non-neutralized the complex tone is still a casual speech phenomenon,

since it has resulted from elision in the first place. The neutralization is contingent on a number of factors, including tempo and register. There is not much point, therefore, in insisting on an exact number of complex tones. Only when the entire business of complex tone is conceived of in this vein can we fully appreciate the picture presented in figure [7].¹

At this point an interesting question suggests itself: to the extent that similar pitch contours, so long as they are not one of the basic tones themselves, are subject to neutralization, how can a complex HR*, whatever its underlying identity, always avoid approximating to T2 and avoid being interpreted or perceived as a T2? The question is especially forceful when the tempo and register favour such approximation and interpretation.

My answer to the question is that the complex HR* does ultimately neutralize with T2.² Thus not only can $ma:j^{5+2}$ and $ma:j^{6+2}$ be neutralized with each other, but each of them can be "coerced"³ into $ma:j^2$.⁴ Theoretically speaking, non-distinction vis-à-vis other complex HR* should not be a prerequisite for a complex HR* to be coerced into T2. Thus, after the initial derivation $ST+1/2 \rightarrow ST+H$ (where T stands for any basic tone), which poses little problem, we can imagine

¹ I am, however, not committed to the details of the hierarchy of neutralization represented there. Among other things, T3 and T5 do not in fact have the same starting pitch (See Section 3.3), and thus should not be collapsed.

² Approximating HR* to T2 is hinted at by Chao (1947:34): 'Words in [T2] never have a corresponding form with [HR*], probably because of the great similarity between this tone and the [HR*]. In fact, a number of cases of the [T2] are really the [HR*] form of some other tone(...) In such cases, the pitch range of the [HR*] form has been shortened and the result is an actual [T2].' We should bear in mind that Chao's stance is [-identical, -variable], and so his HR* includes not only complex HR* but also HR* from T2 switch. Since from our point of view the latter kind of HR* is nothing other than T2, there should be no such a thing as its 'becoming' T2.

³ By coercion I mean the production or perception of an otherwise illformed or ad hoc phonological entity (tone in the present case) as a wellformed and normal one in casual speech and/or the internalization of such a normal lexical representation in the course of lexicalizing an output of a synchronic process.

⁴ The fact that these two particular items are in real life rarely said in T2 (because of the need or intention to avoid ambiguity) is quite another matter. Consider for example laj^{4+2} , which has a high tendency to be realized as laj^2 .

two different courses of derivation towards T2:

$$[14] \quad \begin{matrix} T_i+H \\ T_j+H \end{matrix} \} \rightarrow \text{Figure [7]} \rightarrow \text{Unified HR*} \rightarrow T2$$

$$[15] \quad \begin{matrix} T_i+H \rightarrow \text{NEUTRALIZATION TRAJECTORY}_i \rightarrow \text{HR*}_i \\ T_j+H \rightarrow \text{NEUTRALIZATION TRAJECTORY}_j \rightarrow \text{HR*}_j \end{matrix} \} \rightarrow T2$$

In [14] the various particular complex tones go through the steps of neutralization like the ones depicted in figure [7], whereby a common HR* obtains, which is in turn coerced into T2. In [15] each of the various particular complex tones undergoes some kind of neutralization, after which a complete levelling of all complex HR* does not obtain: there is at best a reduced paradigm of HR*, of which every member may be subject to coercion into T2. Of the two patterns of derivation I contend that [15] is much more likely than [14], as I am very sceptical about the idea of a unified HR* which is not yet a T2.

Earlier we showed that the [+variable] view is more comprehensive than the [-identical, -variable] view, and as such it readily incorporates the latter. Identification of the process of tone coercion, whereby an ad hoc HR* is coerced into T2, a basic tone, seems to serve to bridge the gap between the [+identical] view and the two [-identical] views. This is however an illusion. First, both [-identical] views recognize a unified HR* which is not a T2. Such a HR*, as I suggested, is more imaginary than real, and is not compatible with the [+identical] view. Second, and what is more important, as we have shown in this section, although complex HR* does constitute one source of derived T2, not all derived T2 have complex HR* as their source. This brings us to a major confusion, or failure to discriminate, in the literature between two essentially different kinds of process. One involves the switch from the lower basic tones to T2, and the other involves the elision of the segmental part of a syllable together with other adjustments. The two kinds of phenomenon are different in a number of respects. The following table serves to summarize the differences:

[16]	T2 SWITCH	DESEGMENTED T1 or T2
Is the IA mode applicable?	No	Yes
Is autosegmental representation applicable?	No	Yes
Is the input tone identifiable?	No	Yes
Any intermediate stage of derivation?	No	Yes
Are the processes all productive?	No	Yes
Is the lexicon affected?	Yes	No

Having discriminated these two essentially different kinds of process, we are in a position to give the three views of HR* our final words. None of them discriminates the two kinds of process. Consequently none is an all-round view. The [+identical] view applies well to T2 Switch but cannot handle complex tone. The [-identical, +variable] view applies well to complex tones but fails to capture T2 Switch. The [-identical, -variable] view is intermediate between the two, which does not make it more adequate.

At this point a question remains: given the above formulation of the complex modified tone, where is the border line between HR* and HE*? In particular, should the complex tones T1+T1, T1'+T1, T1+T2, and T1'+T2 and their smoothed forms of various degrees be considered HE* or HR* (or T1 or T2)? These questions cannot be tackled without considering first of all the nature of HE* and T1, to which we now turn.

3.2.2 The high even modified tone and T1

Recall that with regard to the number of contrasting basic tones in Cantonese, the analyses that do not recognize occlusive tones (which usually end up having six tones) involve collapsing the occlusive tones with their plain, non-occlusive counterparts. The collapse works nicely for the pairs T3/T3' and T6/T6'. The collapse of T2 and T2' is even better justified since it enables the process of T2 Switch discussed in the last section to have a unified output. The relation between T1 and T1', on the other hand, is not as straightforward as the other three pairs, for although the HE contour is shared by both T1 and T1', T1 has a HF variant not available to T1'. In any case this difference should not be taken as proof against the collapse of plain vs occlusive tones, given the overwhelming parallel behavior (including, *inter alia*, tone shape) between the two series. Moreover, what difference in realization possibilities there exists between T1 and T1' turns out to be strictly conditioned by the environment: there is a constraint against the occurrence of HF with T1'. The constraint has very plausible phonetic motivation: the occluded syllables are ones that have a voiceless coda; the shorter voiced part is more difficult for contoured tones to realize on.

Now there is the question: apart from the constraint against the occurrence of HF with occlusion, does the occurrence of HE vs HF exhibit any pattern or tendency? Put another way, as far as the plain T1 is concerned, what determines its realization as HE or HF? Different writers give different answers to this question.

For Chao (1947:27), tone sandhi is the only determinant. His account boils down to the following sandhi rule:

[17] T1 → HE / ____ T1,

Zōng (1964) and Y Cheung (1969) maintain a tone-split position with regard to T1. We have argued for the non-split position in Section 4.1.3. Since for neither Zōng nor Y Cheung tone split is the sole

¹ His own formulation is in two rules:

HF + HF → HE + HF

HF + T1' → HE + T1'

determinant for the incidence of HE vs HF, it is desirable for us to take up this issue again.

For Zōng (1964:378-80), an independent tone sandhi process interacts with tone split to determine the occurrence of HE vs HF. Despite the assumed tone split, the sandhi process involved still comes down to the same formulation as for Chao's account.¹ That is to say, in two different parts of formalized rule, the "emic" HE and HF can be collapsed as other writers' T1. The very fact weakens the tone-split position.

Y Cheung's (1969:94-5) account of the distribution of HE and HF is unique and complicated. Like Zōng's account it involves tone split and tone sandhi at the same time, but there is the complication of stylistic variation. Since for Y Cheung a HF does not automatically change to HE before another HF, our sandhi adopted from Chao's account needs one adjustment in order to represent his sandhi, namely changing the environment from T1 to HE, resulting in the following rule:

[18] T1 → HE / ____ HE₂

The sandhi is subject to the complication that T1' (presumably of HE shape) does not constitute a valid environment.² Besides, independently of tone split and sandhi, the stylistic modification "HF+HF → HE+HE" signifies colloquialism.⁴

For Ráo et al (1981:279), the determinant is tone sandhi in association with free/interpersonal/stylistic variation. Their account can be represented by a sandhi rule with optional subparts:

¹ His own formulation is:

$$HF \rightarrow HE / \text{---} \left\{ \begin{array}{c} HE \\ HF \end{array} \right\}$$

$$HF \rightarrow HE / \text{---} T1'$$

In our formulation, while the environment T1 is truly the result of collapsing his HE (including T1') and HF, the left-hand T1 assumes that a left-hand HE takes a free ride on the change T1 → HE.

² His original formulation reads: HF + HE → HE + HE

³ It follows that we have to either treat T1' as emic itself, in compliance with the occlusive-tone analysis, or else align it with HF rather than HE, despite its lack of gradient.

⁴ The form of the rule suggests the effect of tone harmony. We shall not go into it since this is not the focus of our discussion.

[19]

HF → HE / $\begin{cases} \text{HE} <\text{Obligatory}> \\ \text{HF} <\text{Optional, where slow tempo and classicism are} \\ & \text{favourable factors for non-application}> \end{cases}$

One has to say that descriptions of the distribution of HF vs HE are in a state of chaos. Even worse, occlusion, tone sandhi, tone split, and free/interpersonal/stylistic variation have not exhausted all the factors contributing to the choice between HF and HE.

So far in this section we have been dealing with tonal phonemics. It will be seen that tonal morphology also has some bearing on what we have been discussing. Parallel to the case of HR* discussed in the last section, there also exists morphological modification of tone that yields HE*. Again, like the case of HR*, the exact realizational status of HE*, in particular whether it is identical with the HE variant of T1, is one thing that writers do not agree on. Also parallel with the case of HR*, the literature divides into a [+i_dentical] camp where HE* is maintained to be identical with the HE variant of T1, and a [-identic_al] camp where HE* is maintained to be distinguishable from the HE variant of T1.₁

Recall that the complex derived tones brought about by the elision of the segmental part of a syllable include the following cases:

[20] S₁₊₁, S_{1'+1}, S₁₊₂, S_{1'+2}

Recall also that the second element of the complex tone, which may be either T1 or T2, is subject to neutralization as a result of smoothing, ending up as nothing more than a high-pitched end-point, or tail, which can be represented by the letter "H". After the neutralization the maximal four-term contrast reduces to a binary one, namely T1+H vs T1'+H. A reshuffle of these two classes of complex tones yields a binary opposition of a different nature, one that is tone-shape oriented: HE+H

1 Two things are worth mentioning. First, when it comes to HE*, Chao's actual position is not clear: he says that HE 'is almost identical with' HE*, and characterizes both of them as '55' in his own system of tone notation, i.e. HE. Second, unlike the case of HR*, all writers agree that perceptually there is only one realization of HE*.

(from T1+H and T1'+H alike) vs HF+H (from T1+H only).

We have set out to settle the question of how to draw the line between HR* and HE*. Now as far as complex tones are concerned, so long as the first element of the complex is still recognizable, i.e. so long as we can tell apart HE, HF, T2, T3, T4, T5, and T6 as the first element, the question of demarcation is a pseudo-question. It is only at the stage of derivation (in the form of neutralization) when certain (not all) complex tones are coerced into T2 that we begin to worry whether a particular derived tone is T2 or something else. So the proper question to ask is what complex tones can and cannot be ultimately coerced into T2. To this question the answer is that all except HF+H and HE+H can.¹ What, then, do these two complex tones become? After smoothing they become indistinguishable from HE. And this constitutes another source of HE.

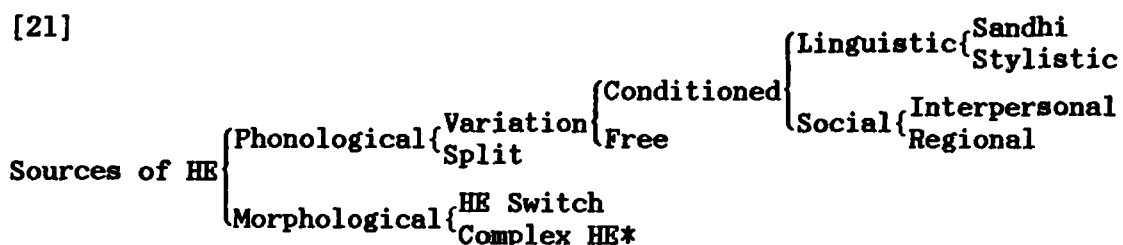
Recall that the process T2 Switch changes a T3/4/5/6 to T2. It is understandable why T2 is not one of the possible input tones. It is not appropriate to speak even of the "vacuous" application of the modification, for since there is no contrast whatsoever (between the "input" T2 and the "output" T2), it follows that nothing can be signified at all. But what about T1? Theoretically it is perfectly possible to switch a T1 to T2, employing the contrast to signify, say, familiarity. Yet the fact of the language is such that T1 is never switched to T2 for any systematic signification. Now, does T1 undergo another kind of modification so as to avail itself of various sorts of meaning signified by T2 Switch? At least for some writers the answer is positive. For those who posit HF as the underlying shape of T1, it is open for them to regard the process "HF → HE" as having the same force of signification as T2 Switch. Hashimoto (1972:182) does exactly this. Though HF and HE are emic for Y Cheung, this does not prevent him from so employing the same process (1969:95-6). HE Switch as an analogue of T2 Switch, then, constitutes yet another source of HE.²

¹ Whitaker's account differs from mine in that her T1* aligns with T2*, T3*, etc. as belonging to her HR*.

² There are limited examples of alternation between T2-5 on the one hand and HE on the other. In view of the small number of items affected and of the fact that the alternation is completely non-productive, lexicalization of the HE version is inevitable. However we treat such minor cases of alternation, the suggested status of HE switch as an analogue of T2 switch is unaffected.

To complete this survey of the factors governing the distribution of HE and HF, it should be mentioned that there is the weakly regionally conditioned variation suggested by Zeng (1982:10). According to him, in addition to the signification of particular meaning or situation of discourse, HE and HF are "sometimes" interchangeable. But he adds that Hong Kong tends to use HE while Guǎngzhōu tends to use HF.

By now we have exhausted the various sources of HE with their respective signifying functions, as reported in the literature. These sources, and their classification, are represented in the following taxonomic tree diagram.



Two questions follow. (i) How can one form, HE, manage to have so many different kinds of content; how can one single signifier signify so many different and potentially competing signifieds at the same time? (ii) In so far as HE has so many different sources, that is, as HE may be the output of so many different processes (in addition to being emic itself for certain writers), the HE must have a very high frequency of occurrence. This is especially the case, and increasingly so, given that the sound pattern of Cantonese has been evolving from an organization favouring an occlusive-tone oriented analysis, where T1 (HF/HE) and T1' (HE only) are separate tonemes, to one disfavouring such an analysis, where T1 incorporates T1'. Given such a situation, how can speakers do otherwise than treat HE as underlying?

The two questions are in fact inter-related, for the answer we provide for the second question nullifies the first: HF is no longer the default value of T1; HE has emerged to be the dominant form, the new default value. Apart from signifying a T1 status, HE effects any other signification only negatively by virtue of being non-HF, while HF, on the other hand, is marked in that it is employed for positive signification.

tion. Admittedly the signified of HF is still complex: HF is still polysemous. But the signified is much simpler than that of the HE in the alternative account depicted in diagram [21]. For me, HF may be a free variant of T1, on which account it signifies nothing; at the same time HF may also signify reading style/formality/solemnity/classicism.¹ For some people there also seems to exist the following sandhi rule:

$$[22] \quad T1 \rightarrow HF / \frac{\quad}{[-occl]} \left\{ \begin{array}{l} \text{pause} \\ \text{low tone} \end{array} \right\}$$

where the demarcation between high vs low tones differs from person to person. In view of the fact that HF and HE are free variants on one plane of analysis, the fluctuating line of demarcation is not unreasonable.²

Treating HE as more basic than HF is no novelty. Despite having "upper falling" as their "suggested name" for T1, Jones and Woo (1912:xiv-xv) remark:

Of the two forms of the 1st Tone the level (...) is by far the commoner. The falling (...) is, however, the normal form at the end of a group, or when a word with the 1st Tone is pronounced by itself. In some cases the level (...) appears to be necessary at the end of a group instead of the falling (...)

This description, made 74 years ago, sounds to me more adequate than any description that explicitly treats HE as basic. Cén (1946:204) gives T1 a single, peculiar shape of "553", i.e. HE followed by a fall, but compares it to the T1 of Mandarin, which is unmistakably of a HE shape.³ Describing Malayan Cantonese, Killingley (1983:3) gives T1

¹ Recall that the optional application of Y Cheung's rule 'HF+HF → HE+HE' signifies colloquialism. Recall also that for Ráo et al the sandhi rule 'HE → HF / HE' is not obligatory in the environment of slow tempo and classicism. In their words, 'in the context of classical expressions non-modification is perhaps preferred.' (1981:279, my translation.)

² Strictly speaking, free variation in association with other conditioned variation is not 'free' any more. At any rate it can be interpreted as having the effect of loosening the otherwise strict conditions of other rules, resulting in what are called *variable rules* in sociolinguistics.

³ Bear in mind that Mandarin has a T4 which is a fall from high to low.

only one value, namely HE, and juxtaposes it against the two values of Kao's T1 in a correspondence table. Rão et al (1981:275-6) are not committed as to the relative dominance of HE vs HF but they note the existence of T1 items that are HE all the time, while the rest of T1 items have alternative values of HE and HF. More recently, Vance's (1977) perception experiment shows that HE and HF alike are perceived as the basic tone T1, and Tse's (1978) child internalizes the T1 as HE not as HF.

Only by visualizing a period of adjustment from T1 as HF to T1 as HE can we understand why descriptions of T1 and HE* are so diverse and appreciate why Zōng and Y Cheung posit the split of T1. Derived HE items lexicalize, just as derived T2 items do. However, whereas lexicalization of T2 items has little bearing on the sound pattern (save for the co-occurrence pattern of tone and occlusion), the accumulative effect of the lexicalization of HE items gradually renders inadequate a phonological account that assumes a T1 to be HF by default. The inadequacy can be remedied to some extent by positing various kinds of process having HE as the output, loading HE with diverse and mythical signification. Split of T1 is certainly compatible with the non-derived nature of HE. It may well be a faithful representation of the state of the language at one stage during the transition from T1 as HF to T1 as HE. As such it is a sound change implemented by way of lexical diffusion, once in progress but subsequently "aborted".¹ As such it could be a faithful representation of the internalized grammar of speakers of a certain age group, especially those who have first acquired a grammar with T1 as HF and are reluctant (not necessarily in a conscious manner) to alter this part of the grammar even when phonological restructuring is called for following an upswing of HE. To the extent that modern linguistics acknowledges personal differences in the grammar internalized by speakers as long as communication does not break down, we see no reason why Zōng and Y Cheung cannot have an idiolect which treats both HE and HF as emic while the majority of speakers choose to switch the default value of T1.

In the light of the discussion in this section, we can draw the following conclusions:

¹ Terminology after Lass (1984:328)

- 1) Complex tones apart, there are no such tones as HE* that are distinguishable from the HE of T1.
- 2) The complex tones T1+T1/2 would ultimately be coerced with HE, but all other complex tones with T2.
- 3) Despite the characterization of T1 as HF in our RD, HE is the unmarked and more frequent value, or default value, of T1.
- 4) HF arises, as a matter of probability, under certain conditions.

3.3 The characterization of tones

Having clarified the inventory of tones, including matters concerning tonal modification, the next question is how we should characterize the six basic tones of Cantonese.

We start from the concrete, phonetic end. Jones and Woo (1912:xiv-xv) use the staff as one of the means to represent Cantonese tones. They note that the characterization represents "average musical value (men's voice)", and add the following comment:

The tones may be transposed into a higher or lower key to suit the voice of the individual student, but their relative values should remain constant. For ladies' voices (average) they should be transposed 8 or 9 notes higher (...)

Now exactly because tones are transposable, the use of the staff to represent them misses the point. Despite using the staff notation, Jones (1912:ix) says, "Students [of Cantonese] who are ignorant of music should go to a singing master, preferably to one accustomed to teach on the Tonic Sol-fa system". He regards the Tonic Sol-fa system as a second choice, suggesting that it is less adequate relative to the staff. Ironically, the notes in the Tonic Sol-fa system are in fact a better simulation of tones than those on the staff, because Tonic Sol-fa is precisely a transposable system of musical scale.

I know of no writer who uses exactly the Tonic Sol-fa system to characterize Cantonese tones. But the system of "tone letters" devised by Chao (1930) comes close to that. Chao (1947:24) explains:

Let the total range be divided into five points (...) A vertical line is drawn as a reference of height and a simplified time-pitch graph is drawn to the left of the reference line. Thus, a sign like 1

stands for a tone which begins high, remains high, and ends high: high level tone.

Chao has simultaneously devised an alternative system, which is isomorphic to the graphical tone letters but uses Arabic figures instead to represent pitch height, from 1 to 5 representing the lowest to the highest pitch in the range. Though he compares these figures to *do*, *re*, *mi*, *fi* and *si* Chao (1947:25) expressly says, "Both the absolute pitch and the size of the intervals depend upon sex, individual, and mood." Thus his systems are at once transposable and relative, as real-life tone systems are. As such they are even more desirable than Tonic Sol-fa for the representation of tones. In comparison, any system of musical scale does not tell us more about a tone as a type, as opposed to (the utterance of) a tone as a token. It is simply inappropriate.

Most works on Cantonese tones use the five-figure system of tone representation. While writers agree on the system of tone representation, they do not seem to agree on how the six individual tones should be represented in this particular system. The following table saves a lot of words.

[23]	T1 T2 T3 T5 T6 T4						FOLLOWERS	
	HK HF	HR	MR	LR	LE	LF LLE		
Cén 1947:204	553	35	33	13	11	221		
Chao 1947:24	55 53	<i>do</i>	<i>do</i>	23	22	21 X	Kao 1971, Dow 1972	
Yuán et al 1960:183	<i>do do</i>	<i>do</i>	<i>do</i>	13	<i>do</i>	<i>do</i> 11	Gão1980, Ráo et al 1981	
S Cheung 1972:5	<i>do do</i>	<i>do</i>	<i>do</i>	<i>do</i>	<i>do</i>	<i>do</i> X		
Hashimoto 1972:92	<i>do do</i>	<i>do</i>	44	24	33	<i>do</i> 22		

The first thing that catches our attention is that for some writers T4 has two possible shapes, despite our RD characterization of T4 as a single LF. The recognition of a very low even (LLE) variant of T4 dates back to Jones and Woo (1912:xv):

The two forms of the 4th Tone may be used indifferently. The level (...) is the easier for Europeans and is therefore recommended. At the end of group it is perhaps safer to use the fall (...)

Yuán et al (1960:188) also give T4 alternative values "21 or 11". So does

Hashimoto (1972:92), though her "even" variant is given as "22" rather than "11" (i.e. not at the bottom of pitch range):

Some speakers prefer the level variant of [T4], probably because of the extreme low register of the low falling contour (...)

Though Fok (1974:12,24) follows Chao 1947 and characterizes T4 as falling, her subsequent more detailed description of T4 suggests that there exists an even variant of T4:

Tone 4 either starts around the level of tone 6 and falls to the bottom of the range or starts at the bottom of the range and stays level. (p.88)

That T4 has a LLE variant, one that is significantly lower than the LE of T6, is a fact of the language. LLE is most likely to occur when the voiced part of the T4 syllable is sufficiently short, either by virtue of the overall brevity of the syllable, as in fast speech or on weak stress, or by virtue of occlusion¹ of the syllable.² A cross (X) in the column headed by LLE indicates that the writer does not recognize this LLE variant of T4.

Cén's representation can easily be dismissed as inadequate. His peculiar representations of T1 (553 only) and T4 (221 only) can be interpreted as hesitation and compromise between the even and falling variants of the two tones, while his giving T6 the shape 11, with average pitch lower than T4, must simply be attributed to a failure in observation.

Leaving aside Cén's inadequacy and some writers' non-recognition of LLF, the differences in the various writers' representation of individual tones become manageable. Thus, after carefully examining the various representations of Cantonese tones, we can infer from them certain cardinal relations (CRs) between the tones, relations which have not so far been violated in any representation except Cén's. The

¹ Refer to Section 7.2.1.1 for T4', i.e. the co-occurrence of T4 with occlusion, so far barely recognized.

² Ráo et al's (1981:276) description is unique: 'The value of [T4] is [LLE] (...), with slight fall in quick tempo, but [LLE] is regarded as the norm in general.' (My translation)

following CRs in particular are significant for understanding the differences in tone representation:

CR-1: T3 and T5 have the same ending pitch.

CR-2: The pitches of HE, T3, T6 respectively and the end-point of LF and LLE are in descending order.

With the help of these two CRs, we can reduce the differences in tone representation to two parametric choices, namely [\pm steep T5] and [\pm extreme LLE]. Thus, since Chao prefers a less steep gradient for T5 to other writers' 13, for him

[\pm steep T5] in conjunction with CR-1 implies that T5=23.

Similarly, since Hashimoto prefers a less extreme representation (i.e. 22) to others' 11 for LLE, CR-2 forces her to characterize T6 as 33 and therefore T3 as 44. And in turn T3 as 44 in conjunction with CR-1 requires her T5 to be 24.

The facility the CRs provide for understanding and predicting the various representations of tones adds to the reality of CRs. It is thus desirable for us to uncover more of these CRs, not only inductively by observing the various representations but also deductively by introspection. Four more follow:

CR-3: T2 and T5 are rising, HF and LF falling, and the rest even.

CR-4: The two variants of T1, HE and HF, start with the same pitch.

CR-5: The two variants of T4, LF and LLE, end with the same pitch.

CR-6: HE and T2 have the same ending pitch.

CR-3 is too obvious to be worth any discussion. Of the other three, only CR-5 is violated, and that by one scholar: Hashimoto (1972:92) has 21 and 22 for T4. The observation of Yuán et al (1960:183) and my own intuition suggest that CR-5 is valid. It is likely that Hashimoto's 22 has to do with her own observation that "phonetically speaking, the Cantonese [T4] starts at a register lower than that of [T6]". She apparently thinks that starting pitch defines T4. In fact what matters is the ending pitch. In the light of the results of an extensive perception study, Fok (1974:88) observes:

Tone 4 either starts around the level of tone 6 and falls to the bottom of the range or starts at the bottom of the range and stays level.

Her detailed acoustic data show that the starting pitch of T4 is not always lower than that of T6 while the ending pitch of T4 is always

significantly lower than that of T6.

This brings us to a kind of consideration that is the central concern of linguistics studies: what differences are significant and what insignificant or coincidental? Though T4 usually starts lower than T6, this is quite irrelevant so long as T4 starts significantly lower than HF and finishes significantly lower than T6. Comparing the starting point of T6 and T4 is an act of irrelevance.

Another irrelevant point concerns the starting pitch of T2 and T5. Despite the constant "3" for T2 and either "1" or "2" for T5 in all representations, the figures seem only to reflect the transcriber's impression that T2 is of higher pitch than T5. But the differing ending pitch is sufficient to tell apart T2 vs T5. Thus Fok (1974:84) observes that both T2 and T5 start at the level of T6. Moreover, her detailed acoustic data show that T2 starts lower than T5 more often than not! This is not really surprising given the fact that T2 is marked, above all, by very steep gradient.

The CRs and considerations of relevance above suggest that none of the representations of Cantonese tones so far is adequate. They also suggest that the differences in representation rest with Chao's scheme of 'description after all. For one thing, the intrinsic four-height pitch contrast in Cantonese tones suggests that a four-height descriptive system would be desirable. Does a five-height system like Chao's tell us more about Cantonese tones than a four-height system? No. Representing a four-height tone system with a descriptive scheme that distinguishes five heights is not unlike notating a three-height vowel system with a four-height notation system (such as the Cardinal Vowel system without using diacritics), when indecision and overdiscrimination are bound to happen.

By translating the five-height representation into a four-height one, taking into account all the CRs and the discussions on relevance, the following representations of individual tones obtain:

[24] HE=44, HF=42, T2=24, T3=33, T5=23, T6=22, LF=21, LLE=11

In this adjusted system of representation the discrepancies between

Chao's [- steep T5] and its opposite and between Hashimoto's [-extreme LLE] and its opposite, i.e. the two parametric choices identified above, are reconciled and therefore no longer exist.

While this representation eliminates much of the irrelevance and incongruence of the former representations, it still suffers from over-distinction. Thus although the set of even contours exhibits a four-height pitch contrast, not all such contrast is exploited by the oblique contours. For example HF can be 41 instead of 42 without being anything but HF.¹ Moreover, HF is in turn just one manifestation of T1, subject to incorporation into the latter. In order to arrive at a more adequate characterization of the individual Cantonese tones, I propose the following binary features:

[*high]: Whether the average pitch is on the high or low side.

[*extreme]: Whether the tone reaches one of the extremes (top or bottom) within the pitch range.²

[*rising]: Whether the pitch rises.

These three features cross-classify the six basic tones:

[25]	T1	T2	T3	T5	T6	T4
high	+	+	+	+	-	-
extreme	+	+	-	-	-	+
rising	-	+	-	+	-	-

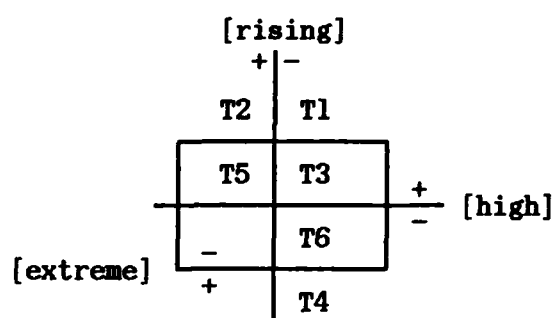
The reality of these oppositions consists in the fact that they interact to define natural classes, which will be utilized in the rest of this

¹ Fok (1974:88) observes, in the light of acoustic data, that 'Tone 1 starts usually at the top of the range and falls to at least the level of tone 6.' See Section 8.2 for details.

² Wang's (1967) system of distinctive features for tones includes [high] and [rising]. In addition to [high], two more features, namely [central] and [mid] also contribute to register contrasts, so that the system is capable of handling five registers. Either one can be used in conjunction with [high] to distinguish four heights. My [extreme] is uniquely defined and as such must not be equated with any of Wang's features, including [central] and [mid]. For one thing, [+extreme] highlights the prominence of T1, T2 and T4, which explains why they are the only tones that can be the output of tone switch. (See Section 7.2.1.1 for T4 Switch.)

thesis. For the time being, we can see how they interact to provide for the general shape of individual tones and classes of tones, as the names of the features imply:

[26]



The three parameters serve to identify only the six basic tones, not the variants of T1 and T4. For the latter purpose an additional parameter [*falling] is needed.

CHAPTER 4: RIME

Recall the following formation rule reproduced from our RD:

[1] $R \rightarrow V (+ Cd)$

It says that the rime is made up of a vowel followed by an optional coda. The present chapter divides into three sections, discussing coda, vowel and the rime as a whole respectively.

4.1 The characterization of codas

Recall the RD array of codas:

[2] Cd = w j
 m n ŋ
 p t k

Following the extraction of the dimension [$*occl$] from segments, the opposition m, n, ŋ vs p, t, k no longer shows up qua codas. Thus the eight-term paradigm of coda reduces to a five-term paradigm, comprising j, w, m/p, n/t, ŋ/k. As shorthand symbols "m, n, ŋ" can of course be used to stand for bundles of properties with the opposition between nasals and stops suppressed.¹ Since there is the syllable-level parameter [$*occl$] already, the usual kinds of segmental distinctive feature that could distinguish nasals from stops, such as [$*nasal$] and [$*sonorant$] are neither necessary nor appropriate here. The mainstream distinctive features [$*continuant$], [$*coronal$] and [$*labial$] then serve well to crossclassify the five contrastive codas:

[3]

	-j	-w	-m	-n	-ŋ
cont	+	+	-	-	-
cor ²	+	-	-	+	-
lab	-	+	+	-	-

¹ Bearing in mind that this does not imply treating the nasals as more basic than the stops.

² The alignment of -j, phonetically a palatal, with -n/t, phonetically dentals, making up a class differing from velars in being [+cor], gives support to the recent change in the conception of

The features interact to define natural classes which participate in various processes and constraints, as the rest of the thesis will show.

4.2 The treatment of vowels

The treatment of vowels is a complicated issue. In this section we first argue for the need to make adjustments to the RD arrangement of vowels, then we provide a critical account of other treatments of vowels, followed in the end by a characterization of the vowels.

4.2.1 Adjustments to Reference Description

Recall the RD array of vowels and the RD configuration of rimes in terms of V+Cd:

[4] V = i:/ɪ y: u:/ʊ
 ɛ:/e œ:/ø ɔ:/o
 ʊ
 a:

[5]

R:	-∅	-j	-w	-m/p	-n/t	-ŋ/k
i:	+	-	+	+	+	[ɪ]
y:	+	-	-	-	+	-
u:	+	+	-	-	+	[ʊ]
ɛ:	+	[e]	+	+	+	+
œ:	+	[ø]	-	-	[ø]	+
ɔ:	+	+	[o]	+	+	+
ʊ	-	+	+	+	+	+
a:	+	+	+	+	+	+

[] = V in variant form
 + = V in basic form
 - = illformed

coronality on the part of some phonologists, who now regard palatal sounds as [+cor]. See, for example, Halle and Stevens 1979 and Halle and Clements 1983.

There are two aspects in which I disagree with the RD presentation above. One concerns the treatment of non-low short vowels, and the other, the relation between [y:] and [u:]. These are discussed in the following sections, followed by a further section presenting the adjusted arrangement of vowels as a result of the discussion.

4.2.1.1 The treatment of non-low short vowels

The status of non-low short vowels ɪ, u, e, ø, o, in particular how each of them is related to the other non-low short vowels on the one hand and to neighbouring long vowels on the other, is a complicated matter. Hashimoto (1972:158) summarises the difficulty involved:

(...) there is more than one possible way of pairing the tense vowels with the lax vowels, or grouping together the lax vowels (...)

The RD treatment as depicted in [4] and [5] represents one way of such pairing and grouping. In essence it consists in recognizing two heights for the non-low short vowels:

[6] ɪ u
 e ø o

and treating each of them as a co-allophone of a neighbouring long vowel in accordance with their roundness, backness and alleged height.

Granted that it gains considerable mileage out of complementary distribution, this treatment, I contend, has a number of drawbacks.

First, if it gets mileage out of complementary distribution *between* short vs long vowels, it at the same time misses the complementary distribution *among* each of two groups of non-low short vowel: ɪ and e on the one hand, and u, o and ø on the other. The two competing cases of complementary distribution are indeed mutually exclusive; a consideration of their relative merits is therefore in order. As Hashimoto (1972:158) observes:

Since there is no natural set of environments that can define the set of lax vowels versus the set of tense vowels there can be no

neat allophonic statement in classical phonetic terms.

But despite such observation, she still "propose[s] to predict the tenseness and laxness of vowels by strictly formulated redundancy statements"(p.158). A similar kind of formulation is also attempted by Light (1977). As things go such rules as formulated for the prediction of vowel length and accompanying change of quality turn out to be complex and unmotivated.¹ In contrast, if ɪ is grouped with e, and u with ø and o, no prediction with respect to vowel length is needed. If there should come the objection that the quality difference between ɪ vs e and between u vs ø and o still needs to be provided for it is time that we corrected the misguided use of phonetic symbols in RD: ɪ is in fact [e] and u, [o]. If anything these two vowels are opener, not closer, than Cardinal Vowels 2 and 7.² While we have settled that ɪ and u are not significantly higher than e, ø and o, the backness difference between ø and o must still be provided for. This can be achieved in two steps. First, the symbol "ø" should also be replaced by "e" in accordance with its actual quality.³ Then, what is more important, the provision for [e] vis-à-vis o and u poses little problem: a fronting rule operates in the motivating environment / — [+cor] (i.e. before -j or -n). The quality of [e] rather than [ø] sets the vowel apart from those intrinsically front, including e which shares with it the same height and shortness.

Second, seen in the light of the actual quality of various vowels in question, assigning e and o respectively to different underlying units from ɪ and u means that predominantly overlapping surface sounds are assigned to different underlying units, and that without good reason.

¹ Neither Hashimoto nor Light recognizes the existence of the rimes ε:w, ε:m, ε:n, om. The omission proves fatal for Hashimoto's 'strictly formulated redundancy statement', which predicts that ε: laxes to e before -n.(p.160) Light's formulation luckily is not affected by the omission but suffers from an empirically wrong (with regard to vowel length) table of rime, which lists ɔ:u for the actual ow and ε:i for the actual ej. As a result his formulation has no provision for ow and ej.

² Refer to Section 8.3 for a full description of the quality of vowels.

³ I admit that there are reasons for preferring 'ø' to 'e'. First, 'ø' is a Cardinal Vowel whereas 'e' is not. Second, 'ø' represents the same height as 'e' and 'o' in a way that 'e' does not. But 'e' represents the quality of the vowel more faithfully, and, being backer than ø, reveals its relatedness to [o].

Extending the argument in terms of distinctive properties, we can see that \emptyset also shares the same height as i and u but is treated as underlyingly of different height, and that again without good reason. The high degree of overlap between the "two heights" of the vowels in questions and the lack of motivation for tearing them apart render such treatment costly and unappealing.

Third, this treatment, as represented in [4] and [5], leads to false predictions. Note the following details of [5], with reminders of the actual quality of the various non-low short vowels:

[7]		$-n$	$-ŋ$
	$\text{æ}:$	$\emptyset[\text{e}]$	$+$
	$i:$	$+$	$i[\text{e}]$
	$u:$	$+$	$u[\text{o}]$

It predicts that:

- (1) $[\text{æ}:n]$ is non-pronounceable, since $\text{æ}:n$ is obligatorily $[\text{en}]$.
- (2) in and un are non-pronounceable, since the environment $/_\text{—}n$ is reserved for $[i:]$ and $[u:]$.
- (3) The non-pronounceable $\text{æ}:n$, in and un , when pressed to be pronounced, will be rendered as $[\text{en}]$, $[i:n]$ and $[u:n]$ respectively.
- (4) $[\text{en}]$ occurs only qua $\text{æ}:n$ and not qua un .

It turns out that none of these predictions are borne out. On the one hand all native speakers of Cantonese can pronounce $[\text{æ}:n]$ without difficulty, as in their rendering of the English sound sequence $/\text{ɜ}:n/$ in such words as "turn", "earn" and "burn": $[\text{æ}:n]$ is never coerced into $[\text{en}]$. On the other hand, a socio-phonological variation where the traditional coda $-ŋ$ realizes as either $[\eta]$ or $[n]$ contradicts all the predictions at once. The variation never involves any change in vowel length. This also applies to the following revealing cases:

[8]	INPUT	OUTPUT
	$\text{æ}:ŋ$	$\text{æ}:n$ ($*\emptyset n[\text{en}]$)
	$iŋ[\text{e}\eta]$	en ($*i:n$)
	$uŋ[\text{o}\eta]$	en ($*u:n$)

The falsification of predictions (1), (2) and (3) is obvious. The

falsification of prediction (4) can be seen in the fact that [ɛn] in fact alternates with uŋ, not with ɛ:ŋ, i.e. it occurs qua un, not qua ɛ:n.

All the foregoing inadequacies are the consequence of treating the non-low short vowels as co-allophones of high and mid long vowels. The problems no longer exist if we amend tables [4] and [5] in such a way that i/e and u/o/e are grouped together respectively and have each group treated as a distinct vowel. The amendment involves non-low vowels only. The relevant part of the amended tables follows:¹

[4'] i: y: u:
e ø/o
ɛ: ɶ: ɔ:

[5']	-∅	-j	-w	-m	-n	-ŋ
i:	+	-	+	+	+	-
y:	+	-	-	-	+	-
u:	+	+	-	-	+	-
e	-	+	-	-	-	+
o	-	[e]	+	+	[e]	+
ɛ:	+	-	+	+	+	+
ɶ:	+	-	-	-	-	+
ɔ:	+	+	-	-	+	+

4.2.1.2 The treatment of y:

According to the RD account of y:, it is a distinct vowel. Since both y: and u: occur in open syllables and before -n, unlike e and o they are not in complementary distribution with respect to the coda. That is to say, as far as the rime is concerned, the vowels y: and u: contrast and their occurrence cannot be predicted with reference to the coda. This explains why they are treated as contrastive vowels on a par with ɛ: and ɔ: in most phonological accounts of Cantonese. Hence our RD treatment of them. However, unlike ɛ: and ɔ:, which are truly contrastive, witness such minimal pairs as gɛ:³ "a saw" vs gɔ:³ (a

¹ In line with the new grouping of the non-low short vowels, i, u and ø will from now on be referred to as e, o and e without any reminder.

classifier) and $g\epsilon:\eta^1$ "ginger" vs $g\omega:\eta^1$ "vessel", it could be maintained that no minimal pair depends on $y:$ vs $u:$, i.e. they are in complementary distribution with respect to the onset. Examine the following table showing the distribution of $u:-$ and $y:-$ -bearing rimes, together with $o\eta$ ("uŋ"), with respect to onsets:

[9]

	b	p	m	f	w	gw	kw	d	t	n	l	dz	ts	s	j	g	k	ŋ	h
y:n								+	+	+	+	+	+	+	+	+	+		+
y:												+	+	+	+				
u:	+				+	+	+	+											
u:n	+	+	+	+	+	+	+												
u:j	+	+	+	+	+	+	+												
oŋ	+	+	+	+				+	+	+	+	+	+	+	+	+	+	+	+

The upper part of the table covers exhaustively the occurrence of the vowels $y:$ and $u:$. It clearly shows the division of labour between them: $u:$ only occurs with labial onsets and $y:$ elsewhere. On the basis of their complementary distribution, Chao (1947) treats $y:$ and $u:$ as co-allophones. Among the twenty-one schemes of romanization and transcription of Cantonese registered in Wu 1976, Chao's scheme is unique in treating $y:$ and $u:$ as non-distinctive. This characteristic treatment by Chao is the topic of much discussion in Kao 1971, Hashimoto 1972 and especially Luke 1983.

Kao (1971:38) treats them as distinct phonemes. She seems to reject Chao's treatment on the ground that the opposition "front" vs "back" is generally distinctive in Cantonese: "it is open to question whether it is justifiable to group [$y:$], which belongs to a front series, with the back [$u:$]". This criticism, however, should not be taken seriously, for we do not require any phonetic property to be "once distinctive, always distinctive".

Hashimoto (1972:156, 164-7) reports an analysis by Shimizu (1963-4:7-16) that also treats $y:$ and $u:$ as non-contrastive, but Hashimoto's discussion is directed to Chao 1947 only. Since she regards Chao's and my $gwu:n$ and $kwu:n$ as $gu:n$ and $ku:n$, she recognizes a problem in Chao's treatment: $u:n$ and $y:n$ contrast after $g-$ and $k-$. ˘

RIME

pays to examine the question in greater detail. As the lenis/fortis pairs of obstruent onsets behave symmetrically, we use K to subsume both g and k (and likewise Kw to subsume both kw and gw).

At the most concrete level of analysis, [Ku:-] and [Kwu:-] can be different. Though Hashimoto observes that [Ku:-] is the actually occurring form, as she says, "Chao's treatment is to consider cases of [u:] occurring after [k], [k'] as derived from cases of [u:] occurring after 'labiovelars'"(p.165), she also notes:

In Jones and Woo 1912, morphemes like ["ancient", "estimate", "drum", "buy"] are given the phonetic transcription of [kwu], which means that, with some speakers at least, the actual pronunciation of certain morphemes coincides with the underlying form given by Chao, although most speakers pronounce these words without the glide. (p.165)

My own pronunciation and observation coincides with the transcription by Jones and Woo, and I doubt if it is the case that "most speakers pronounce these words without the glide".

At a more abstract level of analysis, Kwu:- and Ku:- are not contrastive. As a result [K(w)u:-] can prima facie be interpreted as either Kwu:- or Ku:-.¹ Hashimoto recognizes two disadvantages in treating the sequence as Kwu: rather than Ku:. One follows directly from her recognizing the sequence as phonetically [Ku:]: it costs "a phonological rule, namely, the one that changes sequences of /ũu/ into /u/ if preceded by velar consonants".(p.165)² Another disadvantage is this:

Since homorganic glides are predictable from the following vowels, [ũu:] and [ỹy] need only be represented as /u:/ and /y:/ in the underlying form. However, if these syllables are not distinguished by the feature gravity, then one of them will have to be marked with preceding glide in the underlying form, (...) (p.166)

Weighing the merits and demerits, she concludes, "At present there is no way to judge which of the two underlying configurations is better

¹ It is because of the phonological ambiguity of [K(w)u:-], its compatibility with both Ku:- and Kwu:-, that Luke's survey of whether Kwu: or Ku: is used in 14 dictionaries does not really prove anything.

² Note that Hashimoto treats Kw- as a sequence of a consonant and a glide rather than a unitary onset. This, however, does not affect the cost of a phonological rule.

in the two cases discussed above".(p.166)

Luke 1983 follows the discussion on, and puts forward four arguments in favour of Chao's treatment and against Hashimoto's doubt about it.

First, against the last point made by Hashimoto concerning the sequences [(w)u:] and [(q/j)y:], his reply is that /w/ and /j/ are needed in the inventory of onsets anyway. Though he does not really elaborate the point, I take it that implicit in his reply is the argument that just as underlying u: vs y: without an onset would predict a preceding w vs j, so underlying wu: vs ju: would also predict [wu:] vs [jy:]. While consideration of direction of determination does not favour either treatment, Chao's treatment saves one unit in the inventory of vowels.

Second, the pattern of tense-lax pairing of vowels suggests that [y:] and [u:] belong together:

[10]	TENSE	y:	u:	i:	ɔ:	æ:	ɛ:	a:
	LAX		ʊ	ɪ	ɔ	ə	e	ʌ

While the argument seems convincing on the surface, it is nevertheless not one that Chao would consider. For tense-lax pairing is a position taken by the RD account of non-low vowels, while Chao's treatment of non-low vowels coincides with our amended treatment as shown in [4'] and [5']. Chao's treatment of [y:] and [u:] and his treatment of non-low short vowels seem to be inseparable. We refer again to table [7]. Note that the lower part of the table shows that the distribution of *oŋ* is very different from and much wider than that of either *u:-bearing* or *y:-bearing* rimes. Now if the [o] in [oŋ] (i.e. "u") is treated as deriving from *u:/y:* (which is what the tense-lax pairing is all about), we are faced with the situation where the back allophones [u:] and the [o] in [oŋ] do not occur in parallel environments. As a corollary of the wider distribution of *oŋ* with respect to onsets, the backness of *u:* cannot be provided for simply and generally as "*u:→front / non-labials—*". Rather we have to resort to an additional rule "*u:→lax/—ŋ*", together with extrinsic ordering:

[11]	su:n	bu:j	su:ŋ		su:ŋ
LAXING	—	—	suŋ	FRONTING	*sy:ŋ
FRONTING	sy:n	—	—	LAXING	—

Such a condition, if not impossible, is at least costly. Moreover, we have already given independent motivations for positing emic e and e/o. So argument by appeal to the pattern of tense-lax pairing of vowels is not suitable.

Third, Luke holds that treating the sequence [K(w)u:-] as Ku:- would result in the following skewed distribution of u:-bearing rimes with respect to labial-velar onsets:

[12]	gw-	kw-	w-
	-	-	+

In this shape the argument is not very convincing, for it begs the question of whether w- is needed underlyingly for the phonetic form [(w)u:-]. Those who treat [K(w)u:-] as underlyingly Ku:- (rather than Kwu:-) can always treat [(w)u:-] as u:-, with w- supplied subsequently, resulting in the total non-occurrence of u:- after labial-velars. This, for example, is exactly how Hashimoto handles the data.

Nevertheless Luke is on the right track in recognizing the cost involved in treating [K(w)u:-] as Ku:- with regard to the pattern of distribution, and a change in the detail of this argument would make it more forceful. Instead of aligning f with b, p, and m, I align it with gw, kw, and w, making up a labiodental series, as I have done earlier in table [9].¹ Moreover, the distribution of the u:-bearing rimes with respect to velars is juxtaposed:

[13]	g	k	ŋ	h	gw	kw	w	f
	+	+	-	-	-	-	-	+

Now, even if [(w)u:-] is analysed as u:-, u:-bearing rimes show skewed distribution with respect to both velars and labiodentals. On the other

¹ The question of the alignment of f- with other onsets and the labiodentality of gw, kw and w will be dealt with in Section 5.2.

hand, if [K(w)u:-] and [(w)u:-] are treated as Kwu:- and wu:-, we find a much neater distribution:

[14]	g	k	ŋ	h	gw	kw	w	f
	-	-	-	-	+	+	+	+

Thus, no matter whether the actual pronunciation is [Kwu:-] or [Ku:-], Kwu:- is the preferred phonological shape of the sequence. What is more important, it follows from this treatment that no minimal pair exists contrasting y: and u:, and the neat division of labour between the two vowels, as shown in table [9], is restored.

Fourth, the occurrence of [Ku:-] is symptomatic of an on-going change Kw → K, which is related to the attested socio-phonological variation in the form of Kw → K / — ɔ:.¹ In Luke's own words, the variability of Kw before u: "is no more than an integral part of the change [kw → k] and [k'w → k']"(p.42).² The variation is phonologically conditioned. At any rate it does not occur before low or front vowels. Luke suggests that the environment of the variation might have a wider scope than is generally recognized, including not only ɔ: but also u:.³

Luke's conjecture is well-motivated, in view of the fact that u: and ɔ: share the majority of their distinctive properties, i.e. they fall into a natural class, differing only in [*high]. The lack of actual reporting on the variation before u: may be attributable to the fact that unlike the switch Kw → K / — ɔ:, which constitutes lexical transfer,⁴ in the sense that Kwɔ:- and kɔ:- represent different sets of lexemes, the change Kw → K / — u: constitutes a phonetic drift,⁴ in the sense that Kwu:- and Ku:- are non-contrastive anyway, and the realization of Kwu:- drifts towards that of Ku:-, bringing about more and more overlap.

¹ Following S Cheung 1972, Luke's conception of the variation is such that the rime ɔ: is not included as an environment for the variation. My formulation here describes the situation more faithfully. More on this topic in Chapter 9.

² My translation.

³ Again, Luke's conception of the newly included environment is that it consists of u:n ([*occl]) only, i.e. leaving out the rimes u: and u:j, but as a matter of fact K(w) behaves in the same manner before u: and u:j as before u:n.

⁴ Terminology after Harris 1985.

The point of this argument is that the occurrence of [Ku:-] could well have arisen out of a drift from [Kwu:-] to [Ku:-], meaning that [K(w)u:-] must be viewed as deriving from Kwu:- in order to make the on-going change transparent. Only when pre-u: drift is complete, when Kwu:- has entirely given way to Ku:-, is it desirable to recognize the contrast between Ku:- and Ky:-. Maintaining the underlying Kwu:- "can better reaveal the internal regularity of the sound pattern of Cantonese and the social basis of synchronic variation and historical sound change", otherwise it "will only blur these significant facts of the language".(p.43)¹

To sum up, disregarding his second point concerning the tense-lax pairing of vowels, Luke's arguments, with due adjustment and refinement, serve to dismiss Hashimoto's reservation in favour of the treatment of y: and u: as mutually non-distinctive, which treatment dates back to Chao 1947.

Besides those arguments along Luke's line, the grossly defective distribution of u:-bearing rimes and y:-bearing rimes with respect to onsets also points to the undesirability of treating both vowels as distinct. As will be seen in Chapter 7, there are relatively few co-occurrence restrictions between onsets and rimes. The unusually defective distribution of the u:-bearing and y:-bearing rimes and their mutual complementarity in terms of strict phonological conditioning can hardly be convincingly explained away unless we treat u: and y: as non-distinct.

4.2.1.3 Readjusted arrangement of vowels

In the light of the foregoing discussion, our adjusted arrangement of vowels as shown in [3'] and [4'] needs another round of amendment. The latest version of the table of vowels and the table of rimes follows:

¹ My translation. What I attribute here to the maintenance of the underlying Kwu:- Luke actually attributes to the collapsing of y: and u:. Note that although the latter implies the former, the converse is not true. That is to say, maintaining the underlying Kwu:- is a necessary but not sufficient condition for collapsing y: and u:.

[15] V = i: y/u:
 e e/o
 ɛ: œ: ɔ:
 ɐ
 a:

[16] R:

	-Ø	-j	-w	-m	-n	-ŋ
i:	+	-	+	+	+	-
u:/y:	+	+	-	-	+	-
e	-	+	-	-	(+) ← +	+
o	-	[ə]	+	+	[ə]	+
ɛ:	+	-	+	+	+	+
œ:	+	-	-	-	(+) ← +	+
ɔ:	+	+	-	-	+	+
ɐ	-	+	+	+	+	+
a:	+	+	+	+	+	+

The tables incorporate all adjustments made in Section 4.1.2, including:

- (1) Symbol correction: "ɪ" → "e", "ø" → "e", "u" → "o".
- (2) Arrangement correction:
 - (a) The opposition long vs short is deemed distinctive throughout.
 - (b) ej and eŋ are aligned together, so are ow and oŋ.
 - (c) e and o are treated as co-variants conditioned by the coda.
 - (d) y: and u: are treated as co-variants conditioned by the onset.
- (3) Occurrence refinement: en and œ:n occur qua eŋ and œ:ŋ.

4.2.2 A critical account of other treatments of vowels

We have now arrived at the readjusted arrangement of vowels. There are doubtless other possible arrangements and interpretations of vowels to which I have not devoted any discussion in the course of developing my own version. Hashimoto 1972, for instance, documents nine different analyses of Cantonese vowels. She classifies them into four groups and highlights certain features of each group, sometimes a particular analysis within a group. However, her scheme of classification is, in my view, not entirely adequate, as for example, Wong's (1940) analysis, basically in line with RD in the alignment of non-low short

vowels,¹ is classed with Chao's. She criticizes rather harshly the outlandish analyses of monophthongs as glide + vowel, e.g. [i:] ← je, and of ɤ as a glide, but her evaluation of the various treatments is on the whole not to the point. On certain controversial issues, as opposed to the straightforwardly outlandish treatments, her own preference is not clearly spelt out or rigorously argued for, and can only be inferred (say from her assignment of distinctive features to the various segments). This section aims at giving a critical account of various analyses, drawing on the documentation of the nine analyses in Hashimoto 1972, supplemented with information about the analyses adopted in earlier and more recent works.

The nine analyses in question are those by Wong (1940), Chao (1947), Egerod (1956), Toodoo (1957), Rai (1958), Chén and Bái (1958), Yuán et al (1960), Shimizu (1963-4), and McCoy (1966), documented in Hashimoto 1972:153-6. They vary along the following dimensions:

- [± etic short V]: whether the non-low short vowels are treated as variants of neighbouring long vowels.
- [± etic y:]: whether y: is etic, i.e. a co-variant of u:, or is a distinct vowel.
- [± unitary e/o]: whether e is treated as a co-variant of o, rather than of œ:.
- [± V+glide as diphthong]: whether the phonetic diphthongs are analysed as unitary diphthongs on a par with other vowels, or as V+glide, where the glide is one term in the paradigm of codas.
- [± breaking]: whether ej and ow are interpreted as resulting from the "breaking"² of underlying i:j and u:w respectively.
- [± -jV]: whether certain monophthongs are treated as j+V.
- [± -wV]: whether certain monophthongs are treated as w+V.
- [± etic ε:]: whether a distinct vowel ε: is recognized.
- [± V as glide]: whether certain syllable nuclei are treated as a pre-

¹ Wong's analysis is arguably the most important source of many analyses which form the basis of and popularize the SD and RD treatment of Cantonese vowels, as the Wong 1940 'Syllabary', virtually a pronouncing dictionary, together with its scheme of transcription, has been circulating widely and is received as some kind of norm.

² 'Breaking' is 'the process by which long vowels become diphthongs'. (Sloat et al 1978:117)

vocalic glide, at the same time giving the following high-vowel glide the status of syllable nucleus.

The following table characterizes and classifies the nine analyses in terms of the nine parameters:

[17]	Wong	Chao	Eger	Tood	Rai	Chén	Yuán	Shim	McCoy
etic short V	+	-	+	+	-	-	+	+	+
etic y:	-	+	-	-	-	-	-	+	-
unitary e/o	-	+	-	-	-	-	-	-	-
V+glide as diphthong	+	-	-	-	-	-	-	-	-
breaking	-	-	+	-	-	-	-	*	-
-jV	-	-	-	+	-	-	-	-	-
-wV	-	-	-	-	+	-	-	-	-
etic ε:	-	-	-	-	-	-	-	+	-
V as glide	-	-	-	-	-	-	-	-	+

As the configuration of the table suggests, the first two parameters constitute "major class features", in the sense that they do not mark an idiosyncrasy in a particular analysis as the other seven parameters do. Thus [\pm etic short V] distinguishes two classes of analysis of six and three, while [\pm etic y:] distinguishes two classes of two and seven. This is the reason why I have devoted so much discussion, in fact the entire Section 4.2, to these two parameters. In the course of that discussion, I have also argued for [+ unitary e/o] (though this is a unique option of Chao's), because I regard it as the more satisfactory option. In this regard, then, [\pm unitary e/o] can also be considered a major class feature. Note that [- unitary e/o] has a different meaning according to the value for [\pm etic short V]. For those who opt for [+ etic short V], e must belong with æ: and o with u: or ɔ:, thus implying [- unitary e/o]. A real option of [\pm unitary e/o] is open to those who opt for [- etic short V], when [-unitary e/o] means treating e as distinctive from o despite their complementary distribution.

It is the other six parameters that I have not discussed so far. These six parameters share the characteristic that each of them marks an idiosyncratic treatment, just as [\pm unitary e/o] does, but differ from the latter in that what it marks is what I regard as an inadequate analysis. The critical account that follows will therefore focus on these

six parameters.

[+ V+glide as diphthong] is characteristic of Wong 1940. The first thing to note is that his analysis of Cantonese sounds is by and large an adaptation of the analysis by Jones and Woo (1912). Jones and Woo do not recognize the rime as a phonological unit. Nor do they recognize the fact that the second elements of the phonetic diphthongs [i/y, u] enter into paradigmatic relations with final nasals and stops. My V+glide sequences, therefore, are simply listed as unitary diphthongs on a par with other (monophthongal) vowels rather than analysed on a par with VC sequences. This particular slant is attributable to the influence of the sound pattern of English, either by way of misled views on the part of the analysts or in anticipation of the intuition of potential readers, or both. The diphthong-oriented analysis of Cantonese rimes is quite harmless by itself as supplementary material to a phonetic reader addressed to the English speaking public, but it has far-reaching consequences. Since V+glide is not recognized as such, but rather as a unitary, diphthongal vowel, the first element of the diphthong is not required or expected to coincide with any monophthong. So despite Jones' and Woo's description of the two highest pre-velar vowels as "almost e" and "tend[ing] towards o"(p.xii-xiii) they are free to ignore the relationship between ej and en and between ow and on. Thus, while they transcribe ej and ow as ei and ou, recognizing rather precisely the quality of the first element of the diphthongs, they transcribe en and on as in and un, which is the obvious solution in the organization of the inventory of VC sequences with phonetic diphthongs excluded:¹

[18]		-m	-n	-ŋ
i:	+	+		
e:				+
ɛ:				+
<hr/>				
u:	+	+		
o				+
ɔ:		+	+	

¹ ɛ:m and ɛ:n are omitted in the array of rimes here, assuming the forms had not emerged at the time. Since Jones and Woo do not recognize the rime as a phonological unit, there is no way to know

The influence of the diphthong-oriented analysis by no means ends here. Jones and Woo 1912 is highly regarded by Wong (1940). Though Wong does not hesitate to recognize the Cantonese rime as made up of a V followed by an optional coda where the coda may be high vowels as well as nasal or oral stops (probably because of his acquaintance with indigenous Chinese phonology), he nevertheless adheres to Jones' and Woo's transcription of the rimes in every detail.¹ But transcription presupposes analysis. By transcribing [eŋ] as "iŋ", [oŋ] as "uŋ" and [ɐ] as "ɛ:", he already claims (i) that short vowels are variants of the neighbouring long vowels, and (ii) more specifically that pre-velar [ɐ] and [o] are variants of i: and u: respectively and [ɐ] a variant of ɛ:. It follows that, since he recognizes vocalic codas, the consistent way of handling our ej and ow is align them with either i: and u: (i.e. [+breaking]) or with ɛ: and ɔ:. Yet he takes neither of these options. Retaining Jones and Woo's transcription "ei" and "ou" means waiving the chance of dispensing with the vowels e and o. In light of the foregoing discussion, it can be seen that Wong 1940 basically takes the [+ etic short V] position. The fact that he apparently has not taken this position thoroughly is attributable to the spirit of [+ V+glide as diphthong] that underlies Jones and Woo's transcription, which Wong borrows uncritically. Though Wong's analysis, at least implicitly, is one of [+ V+glide as diphthong] it is more significantly [+ etic short V] in terms of its influence on later writers.

There are other writers who adopt the [+ etic short V] position more thoroughly than Wong, taking either the first option of treating ej and ow as i:j and u:w, as in Egerod 1956, or the second option of treating them as ɛ:j and ɔ:w, as in Toodoo 1957 and indeed our RD. Toodoo is by no means the first person to take that position. As far as I know, Karlgren 1923 takes precedence over all other works in this very treatment. Wang 1936-7 adopts Karlgren's analysis but changes the latter's Swedish dialectological symbols to the IPA. Both Wang 1936-7 and Wong 1940 have become standard references and are thus

whether they accept the sequences ɛ:m and ɛ:n. At any rate this does not affect the point made here.

¹ The only exception is the substitution of 'a' for their 'ɑ:' and 'ɛ' for their 'ɑ'. The alteration is purely notational and bears no consequence or implication whatsoever.

[responsible for the SD/RD treatment in this respect.]

Note that no matter which option is taken, for those who adopt the [+ etic short V] position [eŋ] and [oŋ] have to be aligned with i: and u: for the same reason that leads to the same treatment by Jones and Woo: [eŋ] and [ɛ:ŋ] contrast, so do [oŋ] and [ɔ:ŋ]. The first option aligns [e] and [o] consistently with i: and u:, but the consistency does not extend to [ɛ] anyway: since [en] contrasts with [y:n] (e.g. tsen¹ "spring" vs tsy:n¹ "village"), e is naturally aligned with æ:, a vowel of different height from i: and u:. All [+ etic short V] analyses except Egerod's align [ej] and [ow] with ɛ: and ɔ:.¹ Among them, however, only Yuán et al's (as far as the nine analyses are concerned) is not further marked by any idiosyncratic treatment. This explains why their arrangement of vowels has become a standard treatment, witness our SD/RD.

There is another analysis which does not bear any idiosyncrasy, namely that by Chén and Bái. It is distinguished from others in the configuration of the first three, i.e. the "major class" parameters. Thus, [+ etic short V] distinguishes it from the analysis of Yuán et al, and [- etic y:] and/or [- unitary e/o] distinguishes it from Chao's analysis.

The idiosyncrasy of Toodoo's analysis lies in [+ -jV], which means treating [i:], [æ:] and [y:] as jɛ:, jɔ: and ju: respectively. A number of objections can be raised against this treatment:

- (1) [i:] and [y:] show no sign at all of any opening diphthong.
- (2) Contrasting surface forms would compete for the same underlying form:
 [jɛ:] and [ji:] would compete for jɛ:.
 [jɛ:ŋ] and [jɛŋ] would compete for jɛ:ŋ.²
- (3) A glide -j- is not independently motivated.
- (4) Any glide between the onset and the rime is otherwise unmotivated: it would call for the expansion of the simple canonical shape for

¹ Besides Egerod, Shimizu (1963-4) also aligns [ej] with i:, which is inconsistent because [ow] is aligned with ɔ:. More on this below.

² Admittedly this could be prevented by permitting syllable initial geminate of j. Hence, jɛ: → [ji:]. This expedience, however, is costly. It would call for less restricted canonical shape of the syllable. Moreover, the question of why the syllable jɛ: is not realized as [i:] would still need to be accounted for.

syllable segments "O + R" to a complex one of "O (+ glide) + R".

- (5) The function of the abstract -j- would vary between raising (e.g. jɛ: → [i:]) and fronting (e.g. jɔ: → [æ:]).

Unlike Toodoo, Rai adopts the [- etic short V] position with Chao, but parallel with Toodoo, Rai is idiosyncratic in adopting a [+ -wV] position, which means treating [æ:], [e] (which is a distinct vowel on account of [- etic short V, - unitary e/o]) and [y:] as wɛ:, we (where e is also a distinct vowel) and wi: respectively. Except for the fact that w functions solely to make the following vowel rounded, similar criticism as against Toodoo's [+ -jV] can be voiced against [+ -wV]:

- (1) The treatment is phonetically unmotivated.
- (2) [wi:] and [ɥy:] would compete for wi:.
- (3) The glide -w- is not independently motivated.
- (4) The treatment complicates the syllable formula.

Among all the analyses of Cantonese vowels I have come across, Shimizu's has the fewest distinct vowels: six in all. This is attained by the combination of parameter values [+ etic short V, + etic y:, + etic ɛ:]. I have shown [+ etic short V] to be inadequate. In particular I have shown that it would render the independently motivated position [+ etic y:] less elegant. I believe I have devoted sufficient discussion to [* etic short V] and [* etic y:], and shall say no more about them.

As for the third contribution to the exceedingly small system of vowels, namely [+ etic ɛ:], it is a unique treatment by Shimizu. A prerequisite of [+ etic ɛ:] is that the rimes ɛ:w, ɛ:m and ɛ:n do not exist, which we have shown in Section 2.2.1 to be not the case. Even if we put aside this empirical charge, elimination of the vowel ɛ: can still be shown to be too costly to come by. The arguments follow.

All of the following arrangements conspire to bring about the elimination of ɛ: as a distinct vowel:

- (1) The rime [ɛ:] is aligned with ɐ-bearing rimes, so that [ɐ] is considered the checked version of the [ɛ:] in open syllables.
- (2) [ej] is aligned with i:.
- (3) [ɛ:ŋ] is regarded as alternating with [eŋ] (← iŋ).

[ɛ:], [ej] and [ɛ:ŋ] being the only occurring rimes (assuming the SD rather than the RD state of affairs) that would bear a vowel ɛ:, by

finding a place for each of them Shimizu has actually dispensed with ε : as a distinct vowel. But the arrangements are costly and inadequate in the following respects.

First, the relationship between $[\varepsilon:]$ and $[\varepsilon]$ is not echoed by that between the other half-open long vowels and their short counterparts: compare the $[\alpha:]$ - $[\alpha]$ and $[\ɔ:]$ - $[\mathfrak{o}]$ pairs. $[\varepsilon]$ differs from $[\alpha]$ and $[\mathfrak{o}]$ also in its unrestricted occurrence in checked rimes (Cf. [5]).

Second, aligning $[\text{ej}]$ with i : but $[\text{ow}]$ with \mathfrak{o} : amounts to a partial implementation of [+ breaking]. That is to say, the analogous pair $[\text{ej}]$ and $[\text{ow}]$ are derived differently: $[\text{ej}]$ by breaking (and lowering) but $[\text{ow}]$ by raising.

Third, treating $\text{e}\eta$ and $\varepsilon:\eta$ as alternants takes us beyond phonology into the area of morphophonology, and this treatment does not work even in terms of morphophonology. The morpho-syllables that involve $\text{e}\eta$ or $\varepsilon:\eta$ fall into three classes: (i) $\text{e}\eta$ only, i.e. the morpheme can only be pronounced with $\text{e}\eta$, (ii) $\text{e}\eta$ or $\varepsilon:\eta$, i.e. the morpheme can be pronounced with either $\text{e}\eta$ or $\varepsilon:\eta$, and (iii) $\varepsilon:\eta$ only, i.e. the morpheme can only be pronounced with $\varepsilon:\eta$. A morpheme may exist or not in any one class independently of the existence or not of a counterpart in another class. Discounting the situation where no relevant morpheme exists in any of the three classes, the following seven examples exhaust all the possible patterns of distribution of these three classes of morpho-syllables:

[19]		$\text{e}\eta$ only	$\text{e}\eta/\varepsilon:\eta$	$\varepsilon:\eta$ only
ONSET	T/[$\neq\text{occl}$]			
dz	T6	"silent"	"pure"	(sur $\hat{\text{c}}$ name)
ts	T1	"clear"	"green"	—
d	T3	"knob"	—	"to throw"
ts	T3	—	"red"	"ruler"
l	T6'	"strength"	—	—
ts	T2	—	"to invite"	—
h	T3'	—	—	"to eat"

The very existence of *en*-only and *ɛ:ŋ*-only items, especially when *en*-*ɛ:ŋ* minimal pairs exist, makes the collapsing of [en] and [ɛ:ŋ] impossible.¹

McCoy's analysis, as it is, harbours a three-fold idiosyncrasy:

- (1) [en] aligns with u:.
- (2) [ɐj] and [ɐw] are treated as glide+V.
- (3) [i:w] is treated as glide+u:.

We have captured the last two idiosyncrasies by characterizing the analysis as [+ V as glide], while the first one is left unreflected. Two general remarks on his analysis are in order. In the first place, were his analysis stripped of the said idiosyncrasies, it would be a rather "normal" one, isomorphic with the Yuán et al (or RD) type, who shares with McCoy the first three parameters. On the other hand, each of the idiosyncrasies does not stand and fall with the other two, in the sense that none of them follows from any other.²

¹ Even those morpho-syllables exhibiting an *en*~*ɛ:ŋ* alternation hardly lend themselves to the formulation of any general morphological process. Hashimoto (1972:169-70) gives a list of 38 items that 'exhausts' such alternation. One-third of them have an alternant that I judge unacceptable, e.g. *ben*² 'cake', *ten*¹ 'sitting room' and *lɛ:ŋ*⁶(*ŋɔ:j*⁶) 'besides'. Regarding the truly alternating ones, the relation between the *ɛ:ŋ* and *en* forms is hardly one of neatly [+colloquial] as Hashimoto claims. Many 'alternants' have acquired specialized meaning, and must therefore be deemed to have lexicalized (assuming the alternation did exist), e.g. *den*² 'push against', which even appears in the slang *den*² *lej*⁵*gɔ:*³ *fɛj*³ '(I) push against your lung' despite the expectation [-colloquial] by virtue of the rime *en*. Lexicalization of 'alternants' is hardly surprising in view of the fact that there are *en*-only and *ɛ:ŋ*-only items anyway, which may be [+colloquial] or [-colloquial]. Hashimoto admits that the *ɛ:ŋ*-only items 'must be marked differently in the lexicon from the group that have [ɪŋ]/[ɪk] counterparts' (p.171), but marking must be extended to the *en*-only items and those alternants which has acquired specialized meaning. The need for frequent multiple marking suggests that the relation between *ɛ:ŋ* and *en* as [+colloquial] is not part of the synchronic grammar but a fact of etymology reinforced by common graphemes. The non-productiveness of such alleged relation is good indication of its historical nature. This is not to deny the possibility of an individual making an effort to mnemonically relate any etymologically relatable items as a strategy for the expansion of lexicon, which must be distinguished from the grammar internalized by an idealized native speaker.

² One must be curious to know, if none of the idiosyncrasies follows from any other what kinds of consideration have motivated the idiosyncratic treatments? Since I have access to McCoy's analysis only by way of a table of rimes presented in Hashimoto 1972, together with some sporadic comments therein, I am not in a position to provide an answer. However, judging from the title of McCoy's work, it is probable that he handles the present-day phonological system in such a way

As mentioned earlier, for those who opt for [+ etic short V], the unmarked treatment is to align ej with ɛ: and ow with ɔ:, in line with the alignment of ej with ɛ:. That is exactly what McCoy does as far as ej and ow are concerned. However, while e contrasts with y: and thus cannot be aligned with the latter, it is nevertheless in complementary distribution with u: with respect to onsets: e occurs after non-labials and u: after labials. It is this complementarity that he captures, treating u:n and en as mutually non-distinctive rather than aligning en with ɛ:n, which is the usual course of action given [+ etic short V]. Nevertheless he is not consistent, for ej is still aligned with ɛ:, that is, despite the fact that ej is also in complementary distribution with u:j. The drawback of such treatment is that e is aligned sometimes with ɛ: (ɛ:j → [ej]) and other times with u: (u:n → [en]), where ɛ: and u: differ not only in height but also in backness, and e itself is phonetically closer to ɛ: than to u:. Moreover, while for other [+ etic short V] analyses vowel length is determined rime-internally in general, for McCoy an ad hoc rule is needed to determine the length and quality difference between the u: in u:n and the e in en.

With regard to the second idiosyncrasy, namely treating [ɐj] and [ɐw] as glide+V, Hashimoto (1972:158) comments:

The interpretation of [ɐ] as a glide in the finals /ɛi/ and /ɛu/ (...) contradicts the actual patterning of glides and vowels as captured in generalization that the nondiffuse, nonconsonantal segment is always a vowel, again crippling our general rule for predicting the feature syllabic.

We of course need not follow Hashimoto's descriptive framework, in particular her distinctive feature system, in order to appreciate the undesirability of the interpretation in question. In the first place, as the vowel ɐ and the codas j and w are needed anyway, the leaving of a gap in the slots for ɐj and ɐw needs motivation and costs specific description. Moreover ɐ as a glide does not otherwise exist in the language; positing such an ad hoc glide as ɐ helps explain nothing in

to enable easy link between the 'Proto-Cantonese' he has reconstructed, which is more precisely speaking 'Proto-Yue' (Tsuji 1980:7), and present-day Cantonese. As such his analysis might well represent the phonemic system of an immediate ancestor of present-day Cantonese, which is of value by itself.

the synchronic system. Still more serious is the violation of the canonical shape for syllable segments, "O + V (+ Cd)", which does not allow j or w to exist between the onset and the vowel.

All that is said about əj and əw applies to i:w, which McCoy also treats as glide+V, but one thing differentiates əj and əw from i:w, namely that ə is short while i: is long. If it is grossly counterintuitive to treat the longer part of a phonetic diphthong as a glide, then ə is better qualified than i: to be a glide. Surely considerations of relative sonority between ə vis-à-vis -w and -j and of syllable formula still do not favour the treatment of ə as a glide, and yet such treatment does bear some phonetic plausibility: the ə part of əj and əw is perceptibly shorter than the -j and -w part. This kind of awareness could lead ultimately to the discovery of an important regularity concerning the relative length of vowel and coda in Cantonese, to be discussed in Chapter 6.

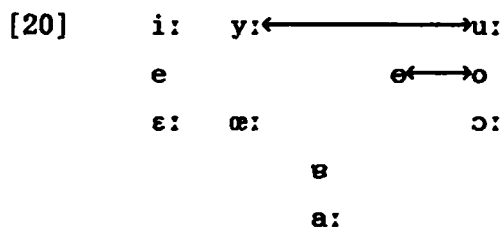
By now a rundown of the nine distinct analyses of Cantonese vowels is complete. I hope to have shown that idiosyncratic, and indeed outlandish, treatments can fairly effortlessly be dismissed as inadequate. They are quite unlike the "major class" parameters, for which a decision can be made only after lengthy and rigorous argument. In the course of such account I have also referred to the analyses by Jones and Woo (1912), Karlgren (1923) and Wáng (1936-7), which predate all nine analyses in question and are jointly responsible for the position [+ etic short V] prevalent today via Wong 1940 and Yuán et al 1960. Apart from these three works which might be considered too old for listing in Hashimoto 1972, there are works that are too recent for it to cover. Among the more recent analyses, Cheng 1968, S Cheung 1972, Gão 1980 and Ráo et al 1981 are demonstrably under the shadow of Yuán et al 1960. Interestingly, despite the fact that Chén and Bái 1958 is "inaccessible to the writer" (p.185), Kao 1971 arrives at the same arrangement of vowels as them.

The position of Hashimoto herself is not crystal clear. As said earlier, she does not spell out her preference, which can only be inferred. This is possible as she adheres to the formalism of (early) generativist phonology, in which the idea of contrasting segments gives way to one of opposing distinctive feature values, and representation-

only description gives way to representation-cum-rule description. For instance, whether or not [y:] is "etic", it has to differ from [u:] in being [-grave]. Anyhow, her handling of distinctive features and discussion of other analyses show that her position is almost identical with Yuán et al: idiosyncrasy-free, and [+ etic short V], which entails [- unitary e/o]. If there is any difference between her and Yuán et al's position, it is that she is not committed as to [* etic y:]: "At present there is no way to judge which (...) is better."(p.166)

4.2.3 The characterization of vowels

The discussion in Sections 4.2.1 and 4.2.2 comes down to the following configuration of Cantonese vowels:



It is clear that the vowels are divisible into sub-systems of long vs short vowels:



Pending further refinement, the long-short difference can be tentatively represented by the SPE tense/lax opposition. Other mainstream generativist phonological distinctive features, in association with [*tense], then interact to cross-crossify the vowels:¹

¹ Compare the use of [*tense], [*diffuse], [*grave] and [*flat] (in that order) in Hashimoto 1972.

[22]	i:	y:	u:	ɛ:	æ:	ɔ:	a:	e	ə	o	ʊ
tense	+	+	+	+	+	+	+	-	-	-	-
high	+	+	+	-	-	-	-	-	-	-	-
low	-	-	-	-	-	-	+	-	-	-	+
back	-	-	+	-	-	+	-	-	-	+	-
rounded	-	+	+	-	+	+	-	-	+	+	-

Owing to the intrinsic characteristic of the feature system, the following implicational relation holds:

$$[23] \quad [+high] \rightarrow [-low] \equiv [+low] \rightarrow [-high]$$

Moreover, the following language-specific implicational relations can be identified:

$$\begin{aligned}
 [24] \quad & [+high] \rightarrow [+tense] \equiv [-tense] \rightarrow [-high] \\
 & [+low] \rightarrow [-round] \equiv [+round] \rightarrow [-low] \\
 & [+back] \rightarrow [+round] \equiv [-round] \rightarrow [-back]
 \end{aligned}$$

4.3 Characterizing the wellformed rime

The latest version of the table of rimes is reproduced below:

[16] R:		-∅	-j	-w	-m	-n	-ŋ
i:	+	-	+	+	+	+	-
u:/y:	+	+	-	-	-	+	-
e	-	+	-	-	-	(+) ← +	+
o	-	[ə]	+	+	+	[ə]	+
ɛ:	+	-	+	+	+	+	+
æ:	+	-	-	-	-	(+) ← +	+
ɔ:	+	+	-	-	-	+	+
ʊ	-	+	+	+	+	+	+
a:	+	+	+	+	+	+	+

It is clear that gaps exist in the occurrence of certain combinations

of V and coda, which gaps can be accounted for in terms of constraints on the combination of V and coda, giving rise to a distinction between wellformed and illformed rimes. The following four constraints are sufficient to filter out all the illformed rimes:

[25]	CONSTRAINT	ILL-FORMED RIMES FILTERED OUT
LAX:	*[-tense]Ø	e, e, o, ɐ
HI:	*[+high]ŋ	i:ŋ, u:ŋ, y:ŋ
YOD:	* $\begin{bmatrix} -\text{back} \\ -\text{low} \\ +\text{tense} \end{bmatrix}_j$	i:j, y:j, ɛ:j, æ:j
LAB:	* $\begin{bmatrix} +\text{round} \\ +\text{tense} \end{bmatrix}_e$ [+labial]	u:w, æ:w, ɔ:w, ew; u:m, æ:m, ɔ:m, em.

Whatever the form of constraint might predict the nonoccurrence of en and æ:n for the majority of speakers, these "gaps" are not accounted for here because they are coming to be filled by variant forms of eŋ and æ:ŋ, which suggests that the apparent gaps are merely historico-accidental.

The occurrence of [e] in place of [o] before certain codas, and that of [y:] in place of [u:] after certain onsets, are matters of realization rather than wellformedness conditions. So are the actual realization of certain codas, especially with regard to lip-shape. In this section we are concerned only with whether a combination is wellformed, not with how the rimes are realized.

The resultant table of rimes, with illformed rimes accounted for with reference to the constraints responsible, looks as follows:

[26]

	-Ø	-j	-w	-m	-n	-ŋ
i:	+	YOD	+	+	+	HI
y:/u:	+	+	LAB	LAB	+	HI
e	LAX	+	LAB	LAB	+	+
ə/o	LAX	+	+	+	+	+
ɛ:	+	YOD	+	+	+	+
æ:	+	YOD	LAB	LAB	+	+
ɔ:	+	+	LAB	LAB	+	+
ʊ	LAX	+	+	+	+	+
a	+	+	+	+	+	+

CHAPTER5: ONSET

In relation to Cantonese onsets, the most interesting questions concern the treatment of glides, labiovelars, and consonant+lateral clusters. These questions will be dealt with in the first section under the heading "clustering". After that we shall attempt a characterization of the system of onsets and compare it with the system of codas.

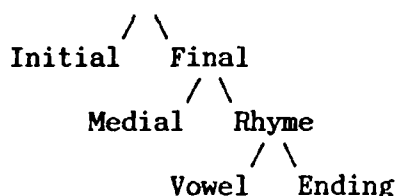
5.1 Clustering

5.1.1 Setting the background

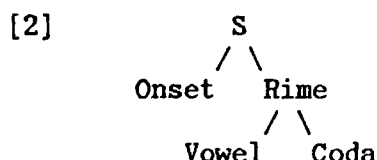
In terms of indigenous Chinese phonology, the segmental component of the syllable of all Chinese dialects takes the following hierarchical canonical form:¹

[1] S=segmental component of a syllable

¹ See, for example, Cheng 1973 and Light 1977. Hashimoto (1972:87-8) also introduces this mode of description, but seems to view the final as flat rather than hierarchical.



Compare the representation used in this thesis:



The correspondence between the sequence "vowel + ending" and "vowel + coda" is obvious. However, because of the difference in the recognition or not of a "medial" in the canonical form of S, one is not in an easy position to equate "rime" with "rhyme", and consequently "onset" with "initial". The non-isomorphism between the two representations is one of the reasons for not adopting the indigenous (but translated) terminology in this thesis.²

Given the syllable structure in [1], one can either say that the "medial" is missing in Cantonese, or else recognize a position for the "medial" in Cantonese. Recognition of the "medial" is not confined to those who adopt the [+ -jV] or [+ -wV] position mentioned in the last chapter. Quite apart from such contrived underlying jV and wV sequences as Toodoo's jɛ: for [i:] and Rai's wɛ: for [æ:], jV and wV exist as normal surface forms, e.g. [ja:] and [wa:], with or without a coda. In the framework of this thesis, these j's and w's are of course onsets. In indigenous Chinese phonology, however, with [1] taken for granted, these j's and w's constitute "medials", not "initials". Syllables beginning with j- or w- are viewed as taking a "zero initial", on a par

² The terminological peculiarity of indigenous Chinese phonology applies to 'initial', 'final', 'medial' and 'ending' only. 'Rime' and 'rhyme', on the other hand, are used interchangeably in the indigenous and other systems alike. Besides, 'Peak' and 'nucleus' are also used interchangeably for what I call 'vowel'. My choice of word is justified on the grounds that only dorsal vocoids appear in the peak/nucleus position in Cantonese. Compare Mandarin for which apical as well as dorsal vocoids appear in this position. The distinction between 'rime' and 'rhyme' is for the sake of ease of exposition.

with syllables beginning with a non-high vowel or glottal stop. This conception of syllable structure has two corollaries. First, no j or w exists as an "initial". Second, Kw- (i.e. kw- or gw-) is a combination of the "initial" K- (i.e. k- or g-) with the "medial" -w-. The conception is characteristic of analyses of Cantonese sounds made with a view to interdialectal comparison, either synchronic or diachronic. Wáng 1936-7 includes an innocent treatment of Cantonese in this vein. In contrast, Chén and Bái (1958:11) expressly characterize their treatment of j- and w- as "medials" as a matter of expedient presentation for the sake of easy comparison with Pekinese sounds.

The question whether Kw- should be viewed as a unitary onset or onset/"initial" plus the "medial" w may be put in another way, one that is not in accordance with indigenous Chinese phonology. There is no principled reason why a non-initial pre-vocalic glide must be a "sister" of the "rhyme" and a "daughter" of the "final", having no direct relationship with the "initial". From an alternative point of view it is the second element in a bi-segmental complex onset, a kind of cluster, with the language-specific condition that only glides occur in this position.³

The "medial" view and the complex onset view of Kw- have different consequences for the analysis of initial glides, as for example in ja: and wa:. The "medial" view regards the glide as a "medial" in a syllable with "zero initial", while the complex onset view regards this initial glide as an ordinary onset. The problem of "zero initial" will be the topic of the next section. As for Kw-, we are not particularly interested in taking sides for either the "medial" view or the complex onset view. These two views can be conflated as the bisegmental view, seeing Kw- as a cluster, in the broadest sense of the term, without any commitment as to the specific hierarchical structure of the syllable. If it can be shown that the bi-segmental view is untenable, there will no

³ Though presented as an 'alternative' point of view, this viewpoint is far from being incompatible with actual organization of the Chinese language (including both its synchronic and historical varieties). [1] is arguably a violent linear (though hierarchical) interpretation of a basically non-linear organization of the Chinese syllable. The medial, being a property of the syllable as a whole, interacts with the 'initial' as well as the 'rhyme'. In Mandarin, for example, the distribution of 'initials' can be best described with reference to the presence or absence of a medial, and which medial is present.

longer be the occasion to waste time on the "medial" vs complex onset issue. The treatment of Kw is the topic of Section 5.1.3.

Proper understanding of the "medial" vs complex onset issue helps us better appreciate the nature of the reported clustering of a consonant followed by a lateral [l] in the initial position (Cl-). Just as Kw- can be either a complex onset or "initial+medial", Cl- can be viewed in either way as well. While [l] is not a glide like the usual Chinese "medials" [j], [w] and [ɥ], all four sounds fall into the natural class of "approximants". Again we are not particularly interested in establishing a characterization of Cl- in either way. Rather we shall ask whether the initial sequence Cl- is needed at all in the underlying representation of Cantonese syllables. This is the topic of Section 5.1.4, followed by another section summarizing the discussion in Section 5.1 on initial clustering in Cantonese in general.

5.1.2 The "zero initial"

In indigenous Chinese phonology a syllable beginning with a glottal stop or dorsal vocoid is deemed to have "zero" as its "initial". This a priori position predicts (1) that the initial glottal stop is allophonically related to initial vowels, and (2) that initial dorsal approximants, e.g. [j] and [w], either form part of the following homorganic high vowel, e.g. [i] in the case of [j] and [u] in the case of [w], or else constitute a "medial", when it is followed by a non-high vowel. Let us look at the facts of Cantonese to see if these predictions are borne out.

As I have mentioned in Section 2.2.2, the onset ɲ- and the lack of onset are no longer mutually distinctive. Even for those who maximally keep these two categories apart, the lack of onset is realized typically as some consonant, usually the glottal stop. The glottal stop does not usually have any allophonic relationship with initial vowels or glides. Vowels (excluding glides) occur initially only either as a result of onset deletion as part of a general process of contraction, or in the limited cases of pre-pausal particles. All this means that the first prediction of the "zero initial" oriented analysis is not borne out in present-day

As regards the second prediction, we observe that the approximants [j], [w] and [ɥ] occur in Cantonese syllables initially, followed by non-high vowels and homorganic high vowels alike. It is clear that these approximants can *prima facie* be treated either as "medials", in which case they are deemed to be preceded by the "zero initial", or as onsets on a par with initial contoids. For either treatment [j] and [ɥ] can be conflated on account of complementary distribution: [j] occurs before unrounded and [ɥ] before rounded vowels. Note that the first treatment eliminates two "initials", j- and w-. In comparison, the second treatment eliminates not only two "medials", but the entity "medial" completely. That is to say, while considerations of inventory size of contrasting segments leave the two treatments equally competitive, the second treatment offers a notably simpler syllable structure. Moreover, the distribution of j- and w- is identical with that of pre-vocalic contoids in that none of them can be preceded by another segment. Treating j- and w- as medials rather than onsets not only would miss the similar distribution of j- and w- vis-à-vis other onsets but also would call for an explanation of why, contrary to the expectation of the framework, they are never preceded by another segment. Since there is no advantage in viewing the initial dorsal approximants as "medials" rather than onsets, the second prediction of the "zero initial" oriented analysis again is not borne out by the facts of present-day Cantonese.

But these arguments are put forward under the assumption that Kw- is a unitary onset, not a sequence of segments. If the bi-segmental view of Kw- is adopted, then at least one "medial", namely -w-, arguably exists. Consequently the entity "medial" cannot be eliminated after all and -w- can be preceded by at least some segments. In contrast, in the light of the foregoing discussion, it is clear that the uni-segmental view of Kw- would be a sufficient condition for the dismissal of the analysis of j- and w- as "medials" following the "zero

⁴ To the extent that accent variation may involve 'systemic' differences (see Section 9.1), it can be argued that for the minority of speakers who maintain a contrast between ɲ- and ʔ-, the ʔ- lends itself to be analysed as a realization of the 'zero initial'. At any rate we still have to examine the second prediction of the 'zero initial' analysis to see if the analysis is tenable.

initial". The exact treatment of Kw-, then, bears crucially on how j- and w- should be conceived, and in turn on whether the idea of "zero initial" is tenable. It is time that we turned to Kw-.

5.1.3 gw- and kw-

In the pronunciation of Kw-, there is simultaneous articulation at both the velar and the labial (either bilabial or labiodental) regions. Yuán et al (1960) argue on these grounds that Kw- is a unitary phonological segment rather than a sequence of two segments. But phonetic evidence of this kind is at best only a weak argument for the status of Kw- as a unitary phonological segment. What is more relevant is the phonological behaviour of the sounds in question. Cén (1946:203), for instance, dismisses the idea of treating the w in Kw- as a medial on the grounds that the w is much more closely related to the preceding consonant than to the rest of the syllable. Presumably, apart from the simultaneous articulation mentioned above, he is also referring to the fact that -w- has a highly restricted combination with the initial sound: it can be preceded by g- or k- only, whereas there is little restriction on what comes after it. The relative lack of restriction on the combination of Kw- with the rest of the syllable also means that Kw- is paradigmatically related to other well motivated onsets.

The existence of the following onomatopoeic items also points to the onset status of Kw-:

- | | | |
|--------|--|------------------------|
| [5] a. | beŋ ⁴ leŋ ¹ ba:ŋ ⁴ la:ŋ ⁴ | "Bang! Bang!" |
| | deŋ ¹ leŋ ¹ da:ŋ ¹ la:ŋ ¹ | "Ding-dong!" |
| | keŋ ¹ leŋ ¹ k(w)a:ŋ ¹ la:ŋ ¹ | "clanging sounds" |
| b. | dzi:ldzi:ldza:ldza:l | "chit-chatting sounds" |
| | wi:lwilwa:lwal | "screaming sounds" |
| | gwi:lgwilgwa:lgwal | "screaming sounds" |

The non-medial status of the w in Kw-, however, does not necessarily mean that Kw- must not be a sequence of two phonological segments. That is to say, though it can be maintained that the w is not a "medial", it is still an open question whether Kw- is a cluster within the system of onsets.

Kao notes that distributionally gw- and kw- "form a class with /w/, rather than with the velar stops /k k'/, in the sense that they impose the same restriction on syllable structure as does /w/". Drawing on the following logically possible combinations of onset and coda which she holds to be "not permissible in the dialect":

$$[6] \quad \begin{pmatrix} w \\ gw \\ kw \end{pmatrix} V \begin{pmatrix} w \\ m \\ p \end{pmatrix}$$

she concludes:

It is clear, then, that the initial /w/ is the restricting element. It will therefore simplify the description if we consider the labio-dorsal stops as clusters, i.e. as /k/ or /k'/ plus /w/.(p.73)

Two objections can be raised against her line of argument. The first has to do with the precise form of the onset-coda combination restriction. There will be a fuller discussion on this topic in Section 7.2.2.2. At the moment it suffices to point out the existence of the onomatopoeic item wow¹ "bark" and the common rendering of the English words "well" as wɛ:w¹ and "cream" as kwi:m¹ by native speakers of Cantonese. Even if the cited restriction in question is accepted as given there is a forceful second argument against treating Kw- as a complex onset. The argument is clearly worded by Lyovin (1973:958):

In the case of affricates, K[ao] rejects the phoneme-cluster solution because it would violate the single-onset canonical shape; so it is at first quite surprising that she eventually interpretes labiovelars as phonemic clusters. (...) Had K[ao] been more conscious of features (...), she could have taken care of the syllable-structure constraint without violating the prevailing canonical pattern of Cantonese. Rather than supposing that it is the presence of the phoneme /w/ which governs the constraint, she could have utilized the feature 'round' which is shared by /w/ and the labiovelar phonemes (interpreted as unit phonemes).

"Forming a class with /w/" is one thing, being treated as clusters of g+w and k+w is another. Sharing some property, say in terms of distinctive features, is sufficient to pull different sounds together "forming a class". The bi-segmental treatment is not the only way, nor

the best way, to capture the distributional similarity. For one thing, -w- would admit only two "preceding" consonants, g- and k-. The same reasons for treating dz- and ts- as unitary onsets apply to Kw-.

Interestingly enough, though Lyovin, without having read Hashimoto 1972, "strongly suspect[s] that it renders K[ao]'s work obsolete", Hashimoto, despite the feature-oriented framework, actually arrives at the same conclusion as Kao on this particular issue:

Among the systematic gaps found in our syllabary as well as among the existent syllable types, none shows any resemblance between labiovelars and velars. (p.138)

[The gaps] all serve to illustrate the patterning resemblance of the labiovelars to the labials, especially to the glide [ʊ]. It is because of this distributional similarity that we consider the labiovelars as combinations of the velar initials with the glide [ʊ]. Then the inexplicable behavior of the "labiovelars" is found to be commonplace indeed. (p.139-40)

This configuration, as already mentioned, will add to existent canonical forms of the syllable (...); but is unavoidable if giving a natural and satisfactory explanation for the behavior of the initials under discussion is more worthwhile than preserving the existent canonical forms of the syllable. (p.140)

If the problem with Kao is that she, as Lyovin suggests, is not conscious enough of distinctive features, the problem with Hashimoto is her pre-occupation with exercising the generativist formalism on Cantonese phonology in terms of rigidly conceived distinctive features. Her argumentation is based on the following feature matrices:

[7]		LABIALS	LABIOVELARS	ʊ	VELARS
Jakobson et al 1952:	diff	+	-	+	-
	grav	+	+	+	+
	flat	-	+	+	-
<hr/>					
SPE:	ant	+	- (~+)	-	-
	round	-	+	+	-
<hr/>					
Wang 1968:	lab	+	+	+	-
	vel	-	+	+	+

She is anxious to class gw-, kw- and w- together in the first place

and to establish a wider class incorporating this class and the labials, but not velars, as a second-order grouping. In terms of the first two feature matrices, the only way out is to dispense with the class labiovelars and treat gw- and kw- as clusters of g/k- and "ũ". Hashimoto seems unaware of the fact that phonological taxonomy and distinctive feature systems are constructs serving descriptions of language and languages, not vice versa. The common practice of regarding "labialvelars/labiovelars"⁵ as velars (primary place of articulation) with lip-rounding (secondary place of articulation) should not be misinterpreted as a claim about the universal nature of any sound with multiple articulation involving the labial and velar regions. In this regard, the comment in SPE:311 is enlightening:

We may ask whether [labiovelars] are labials with extreme velarization or velars with extreme rounding, or, in feature terms, whether they should be represented as (1) or as (2):

(1) $\begin{bmatrix} +\text{anterior} \\ -\text{coronal} \\ +\text{back} \\ +\text{high} \end{bmatrix}$

(2) $\begin{bmatrix} -\text{anterior} \\ -\text{coronal} \\ +\text{back} \\ +\text{high} \\ +\text{round} \end{bmatrix}$

It is clear from the above quotation that SPE is not in fact committed as to the specific configuration of features for what^{are} loosely labelled "labiovelars". Contrary to the conception by Hashimoto, as reflected in the feature matrix (in [7] above) which she attributes to SPE, [+round] is not always a concomitant of what are loosely labelled "labiovelars". SPE cites N.V. Smith's observation that in Nupe "rounded (labialized)

⁵ The terms are spelt or used differently by different writers. 'Thus, we often find the term 'labiovelar' used for either labial-velar (type [p̠k]) or labialized velar (type [kʷ]). In terms of the usual conventions of systematic phonetic terminology "labiovelar" could only mean the anatomically impossible juxtaposition of lower lip and velum.' (Catford 1977:253) Rigorous as Catford's terminology is, a cover term is needed in phonology, if not in phonetics, to subsume labial-velar, labialized velar, and indeed velarized labials. Presumably 'labiovelar' in SPE is used in this broad sense, as subsequent discussion will show. I shall also use 'labiovelar' in this way.

labials are distinguished from nonround labials" and that there are "two types of labiovelars, rounded and unrounded". SPE:311 comments:

The existence of both types immediately resolves the question as to how they are to be represented. We must regard them as labials with extreme velarization (i.e. as having the feature configuration (1), which may or may not also be rounded.

Cantonese exhibits a similar kind of (redundant) distinction between rounded and unrounded onsets, where the lip-shape of the onset assimilates to that of the following vowel.⁶ The regularity applies to non-labials and labials (including gw-, kw- and w-) alike. It follows that Cantonese labiovelars, viewed as unitary onsets, should be regarded as "labials with extreme velarization" rather than "velars with extreme rounding". They fall into a natural class with b- p- m- f- accordingly.

This having been settled, we now turn to the question of how to class labiovelars gw- and kw- together with w-. This is a pseudo-question. Inasmuch as the facts of the language show that w- behaves similarly to gw- and kw-, there is no reason why w- is not just another "labial with extreme velarization". Insisting on the feature configuration of this sound as [+diff, +flat] or [-ant, +round] is a sign of hypostatizing the notation "w" or "ŭ". "w" suggests that it is the functionally consonantal cognate of the qualitatively identical [u].⁷

"ŭ" further suggests that it is not even a consonant.⁸ Phonetically the Cantonese w- often has higher degree of stricture, optionally frictional, in the labial region. Thus not only phonological but also phonetic considerations point to the alignment of w- with unitary labiovelars and labials. There is no need at all to regard labiovelars as combinations of g/k- with -w-, especially when this involves the high cost of complicating the canonical form of the syllable.

⁶ See Section 8.7 for details.

⁷ One source of such suggestion is the IPA scheme of notation. Ironically, the ambiguity of the notation 'j' in the IPA scheme, which can refer to either an approximant or a fricative, might be helpful in guarding against the type of hypostatization under discussion.

⁸ In the Jakobsonian system of distinctive features, there is an added complication resulting from the fact that [diff] is used as a major feature for both vowels and consonants and is interpreted differently in the two cases. High vowels are [+diff] while velar consonants are [-diff]. [w] is thus ambivalent as to [+diff]. The notation 'ŭ' resolves the question in an arbitrary way.

The specific details of onset characterization in terms of distinctive features is not at issue in this section. The spirit of the feature system, I maintain, is that features serve language description, not vice versa. Wang's (1968) features cited in Hashimoto 1972 presented in [7] above, for example, serve well to class together gw-, kw- and w- and in turn these with labials, as Hashimoto recognizes (1972:193). But she dismisses this "way out" without giving any reason. She chooses to conform to the rigidly conceived description formalism of mainstream models, at the expense of descriptive adequacy. The point of my quoting SPE is ^{to} bring home the fact that even for Chomsky and Halle, who are major proponents of description formalism and language universals, considerations of descriptive adequacy still take precedence over the rigid, mechanical exercise of a particular brand of description formalism and compliance with apparent universal principles, which are tentative by nature.

5.1.4 Consonant-lateral clusters

S Cheung 1972:200 documents the following Cantonese loanwords from English:

- | | | |
|--------|---|-------------|
| [8] a. | plej ^l si:2 | ← place |
| b. | ha:jl ^l ka:l ^l si:2 | ← highclass |
| c. | fli:l ^l sa:2 | ← freezer |

Drawing on these three examples, together with sli:m¹ (← slim) from his own observation, Bauer (1984:3) raises the question "[h]ow are these phonetic forms [with Cl-] to be explained and where do they fit into Cantonese phonology ?". Rather than answer this question, one may query the adequacy of S Cheung's and Bauer's observation. Note the different representation of the items in [8] by Y Cheung (1986):

- | | |
|---------|---|
| [8'] a. | pej ^l si:2 |
| b. | ha:jl ^l ka:l ^l si:2 |
| c. | fi:l ^l sa:2 |

There is no mention of the borrowing of "slim" in Y Cheung's article. Note also the following loanwords:

- [9] a. ge⁶li:m¹ ← cream
 b. bət⁶la:n¹dej² ← brandy
 c. kek⁶lek¹dzi:² ← clutch
 d. bek⁶lek¹ ← brake
 e. lɔ:¹ ← gross
 f. si:⁶dɔ:l bɛ:¹lej² ← strawberry (from S Cheung)

- [10] fu:⁶lok¹ ← fluke (from Kiu 1977)

- [11] a. fɔ:l sɔw¹ ← floorshow
 b. fɛ:n¹ ← friend (from Y Cheung)

It is clear from these examples that English loanwords in Cantonese in principle avoid C+liquid clusters. The strategies employed includes C-deletion (as in [9e]), liquid-deletion (as in [8'], [9f] and [11], and rime insertion (as in [9a-d] and [10].

Certainly some kind of "interlanguage"⁹ intermediates between the representation in English and the representation in loanwords. Speakers of Cantonese utter words in their interlanguage as well as using loanwords in Cantonese. Because of the common phenomenon of code switching and code mixing, it is not easy to distinguish with confidence between interlanguage representations and loanword representations. The four unusual forms with syllable-initial Cl clusters identified by Bauer are thus of unclear status. Whether it is plausible at all to treat them as loanwords hinges on whether Cl- clusters exist independently in Cantonese. If Cl- clusters are non-existent in Cantonese, then the Cl- forms under discussion can be categorically dismissed as interlanguage representations.

Cl- clusters do occur, at least phonetically, in a handful of expressions of which the most frequently cited is hɛm⁶bla:ŋ⁶ "all". (See, for example, Hashimoto 1972:19 and Lau 1973:42.) To the extent that Cl-

⁹ 'Interlanguage' is 'the type of language produced by second- and foreign-language learners who are in the process of learning a language.' 'Since the language which the learner produces (...) differs from both the mother tongue and the target language, it is sometimes called an interlanguage, or is said to result from the learner's interlanguage system or approximative system.' (Richards et al 1985:145-6)

clusters do occur, albeit very infrequently, they have to be accounted for and given a place in the organization of Cantonese sounds. As I have mentioned, Cl- can be interpreted either as initial+medial or as complex onsets. I am not going to characterize the sequence in one way or the other; nor am I going to establish Cl- (with various values of C) as unitary onsets. This is because Cl- clusters can all be explained as derived, as the output of phonological processes. A crucial fact about Cl- items is that they all alternate with another form without any Cl-. Ráo et al (1981:295) gives the following examples:

- [12] hɛm⁶b(a:6)la:ŋ⁶ "all"
 dzek⁶b(et¹)lɛt¹ "straight as a ramrod"
 jet¹p(a:j⁴)la:j⁴ "in rows"
 jet¹g(ɛw⁶)lɛw⁶ "in pieces"
 g(ɔ:k³)lɔ:k¹tɛw² "corner"

Bauer (1984:12-3) furnishes the following onomatopoeic examples:

- [13] p(ek¹)lek¹p(a:k¹)la:k¹ "sounds of fire-crackers"
 k(ɛŋ¹)lɛŋ¹k(a:ŋ¹)la:ŋ¹ "clanging sounds"
 p(ɛŋ¹)lɛŋ¹p(a:ŋ¹)la:ŋ¹ "clanging or banging sounds"¹⁰

Bauer observes that "only the labial and velar initial consonants fuse with /-l-/ to form consonant clusters". For this statement to be valid, "labial" has to be construed in such a way as to include labiovelars and labiodentals, as the following onomatopoeic items also exist:

- [14] k(ɛŋ¹)lɛŋ¹kw(a:ŋ¹)la:ŋ¹ "clanging sounds"
 f(i:4)li:lɛ:t⁴ "crying sounds"

Anyway he is right in noting the constraint in terms of place of articulation of pre-l consonants. He adduces the example of [15], which is an onomatopoeic item having similar pattern as [13] and [14]. [16] is my own observation.

- [15] a. dek¹lek¹da:k¹la:k¹ } "ticks"

¹⁰ Bauer also gives the example klek¹kla:k¹, without giving the non-cluster version kek¹lek¹ka:k¹la:k¹, which exists as an alternant of the same expression.

b. d[ə]¹lek¹d[ə]¹la:k¹

- [16] a. si:⁴li:¹sa:⁴la:⁴ } (symbolizing swiftness)
 b. s[ɪ]⁴li:¹s[ə]⁴la:⁴

Note that the (a) forms maximally reduce to (b) forms only, never deriving a Cl- syllable.

Example [16] renders very suspect the "loanword" form sli:m¹ observed by Bauer and cited earlier. In contrast the forms in [8] become more plausible now that we have seen the derivation of Cl- from contraction. If the Cl- and Cr- clusters in English are rendered, by way of rime-insertion mentioned above, as C+rime+l in their loanword forms in Cantonese, the Cl- clusters in [8] can be interpreted as resulting from contraction:

- [17] a. place → pej¹lej¹si:² → plej¹si:²
 b. freezer → fi:¹li:¹sa:² → fli:¹sa:²
 c. highclass → ha:j¹ka:¹la:¹si:² → ha:j¹kla:¹si:²

By now the puzzle of Cl- clusters in Cantonese is solved. They all result from contraction of otherwise normal syllables. As far as the canonical form of the syllable is concerned Cl- clusters have no place at all, in the sense that the restricted inventory of wellformed syllables generated by the syllable formation rules, where only unitary onsets are permitted, is sufficient to account for the existence of Cl- clusters, if a phonological process of contraction in a form hinted at above is recognized. The exact nature of such contraction is not at issue in this chapter, and will be dealt with in Section 10.1.3.

5.1.5 Summary

We have seen that Cl- clusters can be explained as deriving from contraction. As such they have no place in the canonical form of the syllable. It follows that gw- and kw- are the only syllable initial sounds that one might be interested in analysing as bi-segmental. We have also seen that, by analysing gw-, kw- and w- together as "labials with extreme velarization", which can be achieved by whatever feature systems serve the purpose, we have no problem classing them as one

natural class, and in turn these and the labials together as a higher-order natural class. It follows that there is no taxonomic motivation to treat gw- and kw- as C+w. What is even more striking is the paradigmatic relation between g/kw- and other initial contoids. The saving of two unitary onsets, which would be achieved if g/kw- were analysed as bi-segmental, would mean the complication of the canonical form of the syllable. In view of the very small number of initial sounds capable of bi-segmental analysis (dz-, ts-, gw- and kw- exhaustively) it does not pay to complicate the canonical form of the syllable just in order to save a few unitary onsets. This has to do with the highly restricted combination of initial segments if the bi-segmental analysis were adopted. The only possible second segment would be s/z and w, where s/z would only occur after d- and t- and w only after g- and k-. The limited distribution would also need explanation. We can only conclude, therefore, that gw- and kw- are nothing but unitary onsets.

Now since gw- and kw- are unitary onsets, the idea of w-, or of w- and j-, as "medial(s)" also proves untenable. There is no reason whatsoever to complicate the canonical form of the syllable without even saving any entity in the description: it saves nothing to speak of initial j- and w- as "medials" preceded by the "zero initial". Since j- and w- are onsets, not "medials", the idea of the "zero initial" does not apply to syllables with j- or w-. Nor does it apply to syllables beginning with glottal stop, for the glottal stop co-varies with [ŋ-], not usually with vowels. It follows that the whole idea of the "zero initial" is also untenable in present-day Cantonese. By positing an onset ŋ- which incorporates the absence of an onset (which is usually realized as the glottal stop), the onset becomes an obligatory entity in the canonical form of the syllable. It follows, then, that phonologically speaking, onset-less syllables usually only result from onset-deletion as part of a contraction process.

5.2 The characterization of onsets

Now that the unitary status and the inventory of onsets have been determined, we are in a position to characterize the system of onsets. For convenience of exposition I first look at the place and then the manner of articulation of onsets. After that a comparison will be made between onsets and codas.

5.2.1 Place of articulation

The inventory of onsets in our RD is reproduced below, with the place of articulation labelled, as they are traditionally conceived:

[18]	Labials:	b	p	m	f
	Dentals:	d	t	l	
	Alveolars:	dz	ts	s	
	Palatal:		j		
	Velars:	g	k	ŋ	
	Labiovelars:	gw	kw	w	
	Glottals:	(ŋ→)[ʔ]		h	

The distribution of [y:] and [u:] requires that the onsets be dichotomized: labials and labiovelars precede [u:] while the remainder precede [y:]. Then, as we have seen in Section 5.1.4, the labials, labiovelars and velars form a class in constituting the environment for pre-l rime-elision. Since labials and labiovelars are a class already, it follows that the velars themselves form another class.

j-, being the only palatal, can be aligned with alveolars, where there is no sonorant member. The alignment has also a phonetic motivation: the "alveolars" are more often than not realized as alveolo-palatals.¹¹ By the same token, the glottals can be aligned with velars, making up a wider class "gutturals". On the one hand h- fills the empty slot for voiceless fricative the velar class has left behind. On the other hand the variation between [ŋ-] and [ʔ-] is now better provided for as it no longer straddles two places of articulation.

After making the above adjustments, we now have at least three classes of onsets with respect to place of articulation:

[19]	Labials:	b	p	m	f
		gw	kw	w	
	Denti-alveolars:	d	t	l	
		dz	ts	j	s
	Gutturals:	g	k	ŋ	h

¹¹ See Section 8.6 for details.

Though only three classes with respect to place of articulation have so far been justified, the configuration of [19] suggests a system of five places of articulation accommodating at most four onsets each. Realization considerations can be appealed to in order to further divide the larger classes labials and denti-alveolars. Thus, gw-, kw- and w- are marked, not only by velarity as the labelling in [18] suggests, but also by dentality and labial friction. While these properties set them clearly aside from b-, p- and m-, there remains the question whether f- should be aligned with the bi-labials or the "labiovelars". A closer look at the realization of w- resolves the question. Probably because of the predominance of labiality over velarity, the labiality is often realized in the form of labiodentality and/or friction, which marks them off clearly from bilabials, and w- often has no concomitant velar articulation. This suggests that f- should be classed with gw-, kw- and w- rather than with bilabials, making up a class which will from now on be referred to as "labiodentals". More support for this treatment will be furnished when we come to characterize manners of articulation.¹²

Realizational considerations also serve to subdivide the denti-alveolars. Thus, d-, t- and l- are marked by dentality, while the remainder are marked by having the oral stops realized as affricates, i.e. in addition to being non-dental. To accommodate j-, this non-dental series can be labelled "palatals". While s- is not unlike f- in having a manner of articulation for which a contrast between two places of articulation is suspended, the fact that s- is phonetically usually homorganic with dz- and ts- suggests that it belongs to the palatal series. More support for this treatment will be furnished when we come to characterize manners of articulation.

Now that we have established the five-term contrast of place of

¹² The following distributional data are adopted from Hashimoto 1972:139:

	a:w	ɛ:n	ow	i:	i:w	i:n	ɛ:	ej	oŋ	ɞw
b, p, m	+	+	+	+	+	+	+	+	+	+
f	-	-	-	-	-	-	+	+	+	+
gw, kw, w	-	-	-	-	-	-	-	-	-	-

As the table shows, in terms of distribution with respect to rimes, f- is arguably slightly more in line with labiovelars than with bilabials.

articulation it is time for us to characterize the contrast by way of cross-classification in terms of binary distinctive features:

[20]	gw	b	d	dz	g
	kw	p	t	ts	k
	w	m	l	j	ŋ
	f			s	h
coronal	-	-	+	+	-
anterior	+	+	+	-	-
distributed	-	+			

Though [+high] would be needed to describe the actual realization of gw-, kw- and possibly w-, [-dist] is the property that classifies them with f-, resulting in a class occupying one of the five places of articulation.

The treatment of j- as [+cor] echoes that of -j.¹³

5.2.2 Manner of articulation

In the last section we have established a system of five places of articulation accommodating at most four onsets each, with homorganic onsets differing of course in their manner of articulation. The four manners are recognizably (1) lenis/unaspirated oral stops, (2) fortis/aspirated oral stops, (3) sonorants and (4) voiceless fricatives.

The two series of oral stops differ in terms of voice onset time (VOT). As the lenis/unaspirated series have onset of voicing roughly coinciding with stop release (Clumeck 1981) the contrast can be

¹³ See Section 4.1.

captured by either [tense] or [voice].¹⁴ For the time being we leave the choice open.

The oral stops form a class in that unlike the nasal stops they are always voiceless when initial, in addition to being oral and obstruent, and unlike the fricatives they are non-continuant. In the mainstream conception of the feature [continuant] it has to interact with [nasal] or [sonorant] in order to separate oral stops from the rest. In the usage as revised by Dell (1980:xiv,43), nasals are [+cont]. The organization of Cantonese onsets clearly favours this conception of the opposition [±cont]. First, it neatly separates oral stops from the rest. Second, unlike [±nasal], it brings to light the similarity among the "sonorants" m-, ŋ-, l-, w-, j-. And since the "sonorants" are all voiced, we can take advantage of this fact and capture the contrast between these "sonorants" and fricatives in terms not of sonority but of voicing, thus resolving the choice between [tense] and [voice] left open above. [±cont] and [±voice], then, interact to contrast four manners of articulation at each place of articulation:

[21]	gw	kw	w	f
	b	p	m	
	d	t	l	
	dz	ts	j	s
	g	k	ŋ	h
cont	-	-	+	+
voice	+	-	+	-

The following realizational facts give further support to the

¹⁴ Cantonese is actually one of the sample languages used in SPE:327-8 to illustrate the application of the four features voicing, tenseness, glottal constriction and subglottal pressure to VOT phenomena. Based on data from Lisker and Abramson 1964, among the languages in the sample that have only two series of stops in the initial position, Cantonese and English are the only languages where for one series 'onset of voicing substantially coincides with stop release' and for the other series 'onset of voicing lags considerably after stop release'. There is no glottal constriction with Cantonese initial stops. As for [heightened subglottal pressure], it 'is a necessary but not a sufficient condition for aspiration', and coinciding VOT may or may not have [HSP]. One might want to object, in this regard, to the likening of Cantonese to English on the grounds that English lenis obstruents are 'voiced through' between voiced segments, suggesting that Cantonese lenis obstruents never get voiced. But as we shall see in Section 8.5, Cantonese obstruents may also be 'voiced through'.

characterization of the contrast between the third and fourth column in [21] as one of [\pm voice]:

- (1) w- and j- may have friction, resulting in [β] or [v] in the case of w- and [j]¹⁵ in the case of j-.
- (2) w- is often pronounced with dentality, resulting in [ν] or [ν].
- (3) η - is sometimes realized as a voiced pharyngeal fricative [$\var�$].

The glottal stop, another variant of η -, is not as contradictory to the status [\pm voice] as it seems. Though [$\var�$] cannot be "voiced" in the phonetic sense, it echoes the characteristic "coinciding VOT" of the [\pm voice, -cont] onsets. There is no symbol for "voiced glottal stop" for us to employ in order to maintain consistency.

Note also that presence of friction and dentality in w- supports the alignment of f- with gw-, kw-, w-, and presence of friction in j- supports the alignment of s- with dz-, ts-, j-, since w- and j- are the [\pm voice] counterpart of f- and s- respectively in the present scheme.

5.2.3 Summary

The discussion so far in this section (5.2) crystalizes as the following feature matrix:

[22]	gw	kw	w	f	b	p	m	d	t	l	dz	ts	j	s	g	k	η	h
cor	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-	-	-	-
ant	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
dist	-	-	-	-	+	+	+											
cont	-	-	+	+	-	-	+	-	-	+	-	-	+	+	-	-	+	+
voic	+	-	+	-	+	-	+	+	-	+	+	-	+	-	+	-	+	-

To better represent the bundle of features they stand for, "gw kw w" can be written as "gv kv v" and "dz ts s" written as " $j\dot{j}$ cç ç" (or their alveolopalatal/palatoalveolar counterparts) respectively. Notations, however, are merely expedient, short-hand symbols and should never take the place of cross-classifying distinctive features and proper

¹⁵ 'j', not an official IPA symbol, denotes a voiced palatal fricative.

descriptions in the form of both rules and representations. I shall therefore continue using the common roman letters for onsets when phonetic details are not at issue, if only for typographical convenience.

5.3 Comparison with codas

Recall the feature matrix for the system of codas:

[23]	-j	-w	-m	-n	-ŋ
continuant	+	+	-	-	-
coronal	+	-	-	+	-
labial	-	+	+	-	-

A comparison between the system of codas and the system of onsets reveals that the two systems differ in two respects. First, the nasals are [-cont] as codas but [+cont] as onsets. The discrepancy is justified for system-internal reasons. Why *m-* and *ŋ-* are treated as [+cont] has been explained in the last section (5.2.2). Treating *-m*, *-n*, *-ŋ* as [-cont], though in keeping with the mainstream conception of [cont], results in handling apparently identical sounds (*-m=m-*, *-ŋ=ŋ-*) differently, and thus needs some justification. Remember *-m*, *-n* and *-ŋ* are just shorthand symbols for *-[m/p]*, *-[n/t]* and *-[ŋ/k]*,¹⁶ the actual realization of which is contingent upon [**occlusion*]. We need a feature system that treats nasals and oral stops as a class, and the mainstream conception of [cont] serves exactly this purpose.

Second, [lab] is used instead of [ant] for codas. If [ant] were used, the only change needed would be switch the feature value for *-n*, i.e. not as [-lab] but [+ant], and the contrast among the five codas would still hold. [Lab], however, is the more appropriate feature, for neither *-w*, *-m*, *-n*, which would be [+ant], nor *-j*, *-ŋ*, which would be [-ant], constitute a significant class in the language, whereas *-w* and *-m*, which constitute the [+lab] class of codas, belong together in vowel-coda combination restriction.¹⁷

The discrepancy in the exercise of distinctive features between the onsets and the codas brings into focus the fact that they constitute

¹⁶ Glottalization of oral stops is ignored at the moment.

¹⁷ See Section 4.3 for details.

different and independent systems. This illustrates the idea of polysystemicity in the Firthian sense. With polysystemicity borne in mind, we can see that analyses, whether generativist (e.g. Hashimoto 1972) or not (e.g. Kao 1971), which treat the codas as made up of a subset of the onsets, in fact miss the point. This is especially so now that *n-* is no longer a distinct onset. The sheer imbalance in the size of the two systems of contrast should suffice to demonstrate the point: there are 18 onsets but only five (or at most eight) codas. There are doubtless other kinds of evidence too. The two systems, for instance, exhibit different variation patterns: [ʔ] is a variant of *ŋ-* as an onset, but a variant of *-k* as a coda;¹⁸ [n] is a variant of *l-* as an onset but a variant of *-ŋ* in a certain context as a coda.¹⁹ The realization of apparently identical segments is quite different in the two systems too. Thus, *-p*, *-t*, *-k* are glottalized but oral stops as onsets are not. Even *m-* and *-m*, without any notable difference in variation pattern, are arguably differently realized. Because of the difference between "abduction" (for *m-*) and "adduction" (for *-m*) (Saussure 1983:52, 1922:80), it is doubtful if it makes sense to native speakers of Cantonese to say that *m-* and *-m* are the "same" sound. With all this considered, I hold that polysystemicity is firmly established as far as the difference between onsets and codas in Cantonese is concerned.

¹⁸ See Section 8.4 for details.

¹⁹ See Chapter 9 for details.

CHAPTER 6: THE MORA

6.1 Vowel-coda length complementarity

The long-short distinction of vowels in Cantonese might lead one into thinking that rimes, and in turn syllables, also fall into a long class and a short class depending on vowel length. No such claim, however, has ever been made in express terms.¹ If short rimes and syllables are ever recognized, they are considered short by virtue of occlusion, irrespective of intrinsic vowel length.²

This is a rather interesting state of affairs. The vowel, being an integral part of the sequential organization of the syllable (i.e. along the time dimension), contributes to the length of the syllable as a whole. Given the long/short distinction of vowels, either the length difference of the rime (and/or the syllable) has been overlooked by linguists, or there exists some kind of mechanism whereby vowel length difference is offset by the length difference of another part of the syllable.

Something coming close to such a mechanism has been described by Chao (1947:22), but not exactly in terms of length:

An ending is strongly or weakly articulated according as the vowel is short or long. Thus, *an* has a short *a* and a strong *-n*, while *aan* has a long *a* and a weak *-n*.

"Strongly or weakly articulated" is vague description, but it should be clear from the description that there exists some kind of compensation mechanism which serves to maintain a similar level_k^{of} prominence for each

¹ Jones and Woo (1912:xiv) do describe *a:w* and *a:j* as 'long' and *ɛj* and *ɜw* as 'short'. The fact that the long/short distinction applies to just these four complex rimes has to do with two of their other characteristic analyses. First, glide-checked rimes are considered 'diphthongs'. Second, among the monophthongal vowels length is considered contrastive between and only between *a:* and *ɜ* (thus their *a:* vs *a*).

² Cf. Kao 1971 and Lee 1985. As we shall see, this view is misleading, if not categorically wrong.

time as a whole. Put another way, there is complementarity of "strength" between the vowel and the coda.

Chén and Bái (1958) and Yuán et al (1960) give virtually the same description as Chao, except that for Yuán et al strength complementarity applies to contoid codas only. In this regard, Dow (1972:161) observes that "[t]he occlusion of the Cantonese /-n/ following short vowels is more complete than that following long vowels."

The experiments done by Kao (1971) and Lee (1985) enable us to better understand the nature of the complementarity in question.¹

Kao (1971:49) reports the following measurements in milliseconds² of the average vowel and syllable duration for non-occluded checked syllables, i.e. syllables ending in -[m], -[n], -[ŋ],³ -j or -w:

[1]	VOWEL	SYLLABLE
Long V	203	352
Short V	100	294
Long:short ratio	203%	120%

While the long V is twice as long as the short V, the corresponding long syllable exceeds the short syllable by a mere 20%. Assuming that onset length is not affected by vowel length, the length of a sonorant coda is roughly complementary to that of the vowel. The figures show that the sonorant coda is on average significantly longer when the vowel is short than when the vowel is long.⁴ This accords with Chao's description cited above, and what he refers to in terms of articulation strength has correlation with length in the case of sonorant codas.

Note that the figures in [1] are for syllables with a sonorant coda

¹ Neither experiment is concerned with the compensation mechanism in question. The following discussion is mainly my own inference from their data.

² All figures of duration in this chapter are given in milliseconds (ms).

³ 'm, n, ŋ' here represent nasals only, i.e. excluding the homorganic stops.

⁴ I was first brought to the awareness of the long/short difference in Cantonese codas by Dorota Rychlik.

in general. When the coda is confined to vocoids, i.e. -j and -w, the following figures obtain:

[2]	VOWEL	SYLLABLE
Long V	171.5	339
Short V	112.5	308.25
L:S ratio	152%	110%

Though vowel length difference has reduced compared with the figures in [1], length complementarity still holds between the V and the coda.

The similarity between the pattern in [1] and that in [2] shows that Yuán et al's restricting "strength" complementarity to rimes checked by a contoid is an erroneous revision of Chao's description. But if Chao is correct in including rimes checked by a vocoid, is he, together with Chén & Bái and Yuán et al, also correct in including occluded rimes? Consider [3], which is the occluded counterpart of [1], from the same experiment:

[3]	VOWEL	SYLLABLE
Long V	169	207
Short V	89	117
L:S ratio	190%	177%

At the face value of these figures, the extension of the complementarity to occluded rimes encounters difficulty. At any rate compensatory strength of the occlusive coda seems to have no correlate in length, as the long:short ratios of V and the syllable are too close to suggest the existence of length complementarity. In other words, there seems to be no sign that the stop after a short V is significantly longer than that after a long V.

Hashimoto (1972:90) summarises the observation developed thus far:

When the vowel of a syllable is long, the ending is comparatively short, and when the vowel is short, the ending is comparatively long, except when the ending is a stop consonant (...). In general, there is a kind of complementary inter-play between the vowel and the ending.

Close examination, however, reveals that the exclusion of occluded syllables stems from a simplistic interpretation of the figures of syllable length given in [3]. What is involved here is the theoretical issue of how the length of an occluded syllable should be measured. The occlusive codas are simultaneous oral and glottal stops. In these sounds, "the glottal closure, of course, excludes the possibility of pressure build-up behind the oral stop, which may thus be barely audible, except as a distinctive 'on-glide' and 'off-glide' to and from the glottal stop." (Catford 1977:190) In the case of Cantonese the hold phase of a glottalized stop, which is functionally a coda, lingers for a moment, and the release is never audible, whether or not it is followed by another syllable. As such, the occlusive coda carries no acoustic signal, save for the on-glide towards the stop. To the extent that Kao's measurement is done on acoustic signals, the duration of the occlusive codas must turn out to be practically zero. One might want to say that if that be the case, then one has to admit the lack of duration of occlusive codas and the resultant shortness of occluded syllables. According to this view, the long/short distinction in non-occlusive codas applies vacuously to occlusive codas, which have no perceived duration whatsoever. I argue that this is a simplistic, and therefore inadequate view of coda length and syllable length, for there are reasons to believe that it does make sense to speak of the length of occlusive codas.

First, in connected speech, or whenever an occluded syllable is followed by another syllable, the second syllable does not start exactly when the on-glide to the occlusion finishes and the hold phase is reached. There is a time lag between this point and the point when acoustic signals of the following syllable begin to show up. This period of time when there is no acoustic signal must be assigned to the first syllable and deemed to represent the duration of the occlusive coda, which adds to the duration of the occluded syllable.

Our second argument follows from, and therefore presumes, syllable isochrony. A fuller argument for syllable isochrony will be given in Section 6.3. At the moment we are concerned with extending the otherwise reasonable notion of syllable isochrony to the apparently characteristically short occluded syllables. Consider the average duration, again in milliseconds, of three types of syllables reported in

Kao 1971, as displayed in [4]:

[4]	SYLLABLE TYPE	DURATION
	(a) Unoccluded	327
	(b) Long V occluded	207
	(c) Short V occluded	117

Ignore the rough duration ratio 3:2:1 at the moment, the significance of which will be clarified as we move on. Despite the marked discrepancy in these figures, a syllable of any type has the same possibility of assuming a beat in connected speech. If the intuition of myself as a native speaker may not be considered sufficiently reliable, Jones and Woo's (1912) meticulous representation of "rhythm" in terms of syllable length throughout the text of Cantonese conversations should serve as convincing evidence. In their text, while the medium length syllable prevails, there are occasional shorter and longer syllables; but the three lengths are independent of the syllable types as identified in [4], and in fact cut across those three types. Not only do type (b) and type (c) have the same rhythmic tendency and potential, but there is no difference in this regard as to whether the syllable is occluded or not. The phenomenon is echoed in the recitation of classical Chinese verse in the pronunciation of present-day Cantonese, where the three types of syllable receive identical rhythmic treatment. This can only be possible if the inaudible (but articulated) part of the syllable, where the occlusive coda lies, counts as contributing to the whole length of the syllable.

Nothing prevents the mute hold-phase of a stop from functioning positively and assuming a structurally significant position in speech and in verse recitation. Just because it is the prolongable hold-phase rather than the momentary release which realizes an occlusive coda, an occlusive coda lends itself to analysis on a par with a non-occlusive coda with respect to length, so that length complementarity applies to occluded as well as un-occluded rimes.

To recapitulate, Kao's figures strongly suggest the working of length complementarity between the V and the non-occlusive coda. Her figures seem to rule out such complementarity in occluded rimes, but we have seen that this interpretation stems from an inadequate view in

relation to the duration of occlusive codas and therefore of occluded rimes and syllables. Though we do not have corroborative figures supporting length complementarity between the V and the occlusive coda, we hypothesize that the same mechanism of length complementarity works for all kinds of checked rimes on the basis of such indirect evidence as syllable isochrony.

6.2 A moraic interpretation of vowel and coda length

The next question is, how should the regularity of length complementarity be formulated? The nature of the regularity suggests that there exists some relatively fixed measure of length, which is to be divided between V and Coda. A unit of length in language is customarily referred to as a "mora". In order for the length of vowel and coda to vary there should be at least two morae for the V and coda to compete for. The autosegmental framework of phonological representation, with the possibility of more than one tier of discrete units, permitting one-to-many association of units in different tiers, makes possible an elegant description of the length complementarity in question. Thus, let M be mora and Cd be coda; the two kinds of vowel-coda configuration can be represented as [5]:

[5] a.	V Cd	b.	V Cd
	\		
	M M		M M

[5a] represents a long V, short coda configuration and [5b], short V, long coda.¹ [5] describes and explains (i) that vowels and codas alike exhibit a long/short distinction, i.e. [\pm long], and (ii) that tauto-rimic vowel and coda have opposite values for the variable [\pm long]. On top of this [5] also predicts that checked rimes have by and large equal length. In order for the mora to have greater explanatory power we also expect not only checked but also unchecked rimes to be bi-moraic. In other words we also expect unchecked rimes to have by and large

¹ V Cd is another logical possibility to represent the short V,
 | / |
 M M

long coda configuration instead of [5b]. My choice will be justified later in this chapter.

equal length to checked rimes.

No measurement has ever been made on the rime (at least not in its capacity as a rime). So we can tackle the question of rime length only indirectly. Consider the average V length in different rime types as reported in Kao 1971:49.¹

[6]	Unchecked	308 (a)	
	Checked	$\left\{ \begin{array}{l} \text{Long } V \begin{cases} [-occl] \\ [+occl] \end{cases} \begin{matrix} 203 \\ 169 \end{matrix} \\ \text{Short } V \begin{cases} [-occl] \\ [+occl] \end{cases} \begin{matrix} 100 \\ 89 \end{matrix} \end{array} \right\}$	average 186 (b) average 94.5 (c)

Ignoring the insignificant vowel length difference governed by occlusion, we can see that the length ratio (a):(b):(c) is roughly 3:2:1. The same ratio also obtains in a similar measurement of vowel length conducted by Lee (1985):²

[7]	Unchecked	280	
	Checked	$\left\{ \begin{array}{l} \text{Long } V \begin{cases} \text{nasal coda} \\ [+occl] \end{cases} \begin{matrix} 196 \\ 158 \end{matrix} \\ \text{Short } V \begin{cases} \text{nasal coda} \\ [+occl] \end{cases} \begin{matrix} 99 \\ 83 \end{matrix} \end{array} \right\}$	average 178 average 91

Remember that the figures in [4] also exhibit the ratio 3:2:1. This is not surprising in view of the fact that the occlusive coda is there deemed to have only negligible duration and of the assumption that the duration variation of the onset is insignificant. That is to say, the only difference between the figures in [4] and the corresponding figures in [6] and [7] is that [4] includes the onset. The same ratio, then, obtains in different experiments and perspectives. Note that while no measurement of the rime qua rime has been taken, rime duration coincides with V duration in the case of the unchecked rime. Although the ratio does not directly show isochrony between checked and unchecked rimes, the extra duration regularly attached to the

¹ The average figures [6b] and [6c] are not given in Kao's report. Nor are the corresponding averages in [7] given in Lee's report.

² Note that rime checked by glides (-j, -w) are not included in Lee's report.

unchecked vowel can only be explained by the vowel's capacity as a rime, and thus corroborates the isochronous rime hypothesis. We envisage the autosegmental moraic representation of the unchecked rime as [8]:



The fact that the second mora, unlike that in [5a], is solely linked to the vowel explains why the vowel is longer than the checked long V.

Neither Kao nor Lee mentions or shows any awareness of the 3:2:1 ratio. They follow the traditional line of making a binary distinction of long and short vowels, lumping together the unchecked V and the checked long V as a single category of "long vowels". Thus, the conclusion Lee draws from Kao's measurements in [6] is that the long:short ratio of vowels is 226:95, where 226 is the average of the first three figures (i.e. 308, 203 and 169). The conclusions Lee draws from his own measurements are:

(a) The five-grade hierarchy of vowel length in [7] replicates that in [6].

(b) The duration-range of the long V and that of the short V basically do not overlap as far as the individual speaker is concerned.

While (a) is observationally correct, the five figures (with the two average figures furnished by me excluded) in [6] and their counterparts in [7] on the one hand fail to capture the striking and theoretically inspiring duration ratio 3:2:1, and on the other fail to recognize that the differing effects on vowel duration of occlusive vs non-occlusive codas are an independent phenomenon of secondary significance compared with the extra length attached to unchecked vowels and indeed with the 3:2:1 ratio. Again, while (b) is also observationally correct, it is a pity that the regular length difference between the unchecked V and the checked long V, which suggests rime isochrony, is played down.

The three-way length difference of vowels also accords well with the following observation by Hashimoto (1972:90):

[T]here are in fact three gradations of relative phonetic length among the vowels (...): longer when the ending is zero, long when the ending is short, and short when the ending is long.

Shall we then speak of three classes of vowels with respect to length? No. I maintain that the traditional dichotomy of vowels into a long class and a short class, with the unchecked vowel belonging to the long class, is in a broader perspective well motivated. The motivation is three-fold.

First, the long/short distinction has a concomitant qualitative difference. Thus shortness and the qualities [e e o e] are concomitant characteristics. These qualities and the unchecked environment are mutually exclusive. On the other hand the qualities that are concomitant with (longer) length in the checked environment, namely [i y u ε æ ɔ a], are exactly the only ones compatible with the unchecked environment.

Second, as the unchecked vowels are the longest of all, they lend themselves to conflation with the checked long vowel: in both cases the V is longer than the short V (which is necessarily checked).

Third, in terms of autosegmental moraic representation, the unchecked V and the checked long V share the characteristic that the V is linked not only to the first but also to the second mora, i.e. it is ambi-moraic (Cf. [5a] and [8]), whereas the (checked) short vowel is mono-moraic (Cf. [5b]). Whether the second mora is shared, as in [5a], or monopolized by V, as in [8], then constitutes a secondary difference. The moraic representation not only serves as an independent motivation for classing the unchecked V with the checked long V, it actually accounts for the regularities described in the last two paragraphs.

For these reasons the long/short dichotomy of vowels, which we have captured in feature terms tentatively as [**tense*], should still be maintained despite the three-way length difference of vowels as revealed in the measurements cited.

6.3 The mora and syllable isochrony

Length complementarity between V and coda, the extra length

attached to unchecked vowels, and the 3:2:1 ratio all suggest rime isochrony. The moraic representation of the rime even predicts rime isochrony. However, rime isochrony is so far only accepted on the basis of two assumptions, namely (i) syllable isochrony, and (ii) insignificant length variation of the onset, which can be rephrased as simply "onset isochrony". Here we first try to justify syllable isochrony, for which there is independent motivation.

One line of argument is by appeal to native speakers' feeling for the rhythmic nature of successive syllables. Jones and Woo 1912 is probably the only work that makes explicit reference to, and meticulous transcription of, Cantonese rhythm. Thus Jones (1912:vii-ix) writes:

The rhythm which is such a characteristic feature of Chinese pronunciation is indicated throughout this book. In the texts accompanied by musical notes it is shown by the length values of the notes (crotchets, quavers, etc.); in the other texts the lengthened syllables are given in thick type and the very short syllables in italics.

Though four length values of musical notes are used in the staffed text, namely crotchets, quavers, semiquavers and demisemiquavers, only three kinds of syllable lengths are in fact recognized: syllables with a contour tone are as a rule represented by a sequence of two notes, each note half as long as the even-toned syllable. The same rhythmic distinctions are maintained in the non-staffed "other texts". Syllable isochrony can be seen in the fact that the overwhelming majority of the syllables are represented by either a quaver or two semiquavers, corresponding to the medium roman typeface, as opposed to thick/bold type (double-length) and *italics* (half-length) in the non-staffed texts. The occasional shortening and lengthening of a syllable should not constitute any counterargument against syllable isochrony. What I intend by "syllable isochrony" is that the internal structure and constituent elements of the syllable have no effect whatsoever on its "length", conceived not in terms of the presence or absence of acoustic signals but of its tendency and potential with regard to all kinds of rhythmic treatment. Just because syllable length plays no role in the identification of the structure and constituents of the syllable, and because of the clear tendency for syllable lengths to converge at some neutral value, the possibility exists for the speaker to manipulate syllable length by departing from this neutral value to achieve special

effects both in natural speech and in verse recitation. What the three-way distinction of syllable length in Jones and Woo 1912 shows, for instance, is the exploitation of syllable length, or more exactly syllable rhythm, to signify different degrees of sentence stress. In particular lengthening signifies emphasis and shortening signifies insignificance. The same reasons explain the possibility of exploiting syllable length variation for artistic effect in verse recitation.

Structure-independence of syllable length is seen in the fact that syllable length variation, both in natural speech as that given in Jones and Woo 1912 and in regulated recitation of verse, cuts across all types of syllables. That is to say, though variation occurs, the variation is not correlated with vowel length or coda type.

Apart from the impressionistic account of syllable isochrony given above, Kao's measurements of syllable duration also suggest syllable isochrony. Consider the duration of the vowel and syllable in different types of syllables:

[9]		VOWEL	SYLLABLE
	Unchecked	308	335
	Checked, unoccluded		
	{ Long V	203	352
	{ Short V	100	294

Because of the problems involved in the interpretation of the length of occluded syllables, they are not considered for our purpose here. [9] shows that despite the marked difference in V length the difference in syllable length is very small, and can therefore be disregarded.

On the basis of the various kinds of evidence given above, we take the case of syllable isochrony as established, independently of any consideration in terms of the mora.

As for onset isochrony, the assumption is based not so much on published descriptions or hints from acoustic measurements as on the intuition of native speakers, and in fact despite apparent indication to the contrary arising from acoustic measurements. Though neither Kao nor Lee makes any direct measurement on the duration of onsets, the measurements of vowel duration and syllable duration in the case of open syllables give indirect figures of onset duration. The figures so

obtained range from 2 ms for hɔ:1 "exacting" and 100 ms for hæ:1 "boots". The duration of non-sonorants, however, is notoriously difficult to determine on the basis of acoustic signals. For instance, formant transitions, which are the most important clue for the identification of particular stops, may be interpreted as belonging either to the stop or to a neighbouring sonorant. Despite the wide range of values to be observed, the intuition of native speakers points to the insignificance of onset length variation. Compare V length difference, which is well above the level of consciousness. With proper orientation two discrete lengths of coda can be felt too. In contrast no comparable difference in onset length is regularly maintained or even detected impressionistically.

The problem of onset isochrony can be best tackled if considered together with syllable isochrony and moraic organization. In moraic terms syllable isochrony means that the number of morae co-extensive with the syllable is fixed. We have hypothesized that the appropriation of two morae between V and coda accounts for length complementarity. I propose that we do not need any additional mora to represent syllable length. That is to say, the syllable is co-extensive with two morae. Given the configurations in [5] and [8], the onset is always linked to the first mora:

[10]	a.	O V	b.	O V Cd	c.	O V Cd
		/		/ \		/
		M M		M M		M M

The representations in [10] at once explain the following things:

- (1) Syllable isochrony: this follows from the fact that every syllable is linked to two morae.
- (2) Rime isochrony: this follows from the fact that every rime is linked to one mora and a half.
- (3) Onset isochrony: this follows from the fact that every onset is linked to a mora shared also by the vowel, i.e. the onset is semi-moraic. The fact that the upper limit of onset length in Kao's examples is 100 ms or a little less than one quarter of the syllable length, also corroborates [10].
- (4) Threshold of length awareness: since the lower limit of length contrast exercised in the language is, I would claim, a difference of

a semi-mora, the semi-mora can be said to be the threshold of length awareness in the language. Despite the high degree of deviation in onset length, since and as long as the variation is below the semi-moraic level speakers are not aware of it.

- (5) The 3:2:1 ratio of V length: this follows from the fact that the respective vowels are sesqui-moraic ([10a]), mono-moraic/bi-semimoraic ([10b]) and semi-moraic ([10c]).

The moraic arrangements in [10] also resolve a possible indecision as to whether the rime with short V should be represented as [11a] or [11b].

[11]	a.	V Cd	b.	V Cd
		/		
		M M		M M

Though V-coda length complementarity can be accounted for in either representation, [11a] would lead to the replacement of [10c] by [12].

[12]	O V Cd
	\ /
	M M

This would have the following consequences:

- (1) We would be drawn below the semi-mora level of length characterization, which is otherwise avoidable.
- (2) There would be no motivated relation between the 3:2:1 V-length ratio and the moraic structure.

Since these consequences are undesirable, I dismiss [11a] in favour of [11b], thus preserving [10].

8.4 A moraic characterization of rime types

We have seen in Chapter Four that writers who are not fully aware of the paradigmatic relationship between syllable-final vocoids (i.e. -j and -w) and syllable-final contoids tend to characterize rimes checked by -j or -w as diphthongs, as these rimes phonetically are. This is what Jones and Woo (1912), for example, do. In analogy to the long/short distinction between the monophthongs a: and ə, however,

they describe a:j and a:w as "long" and ej and ew as "short". In the light of the foregoing discussion in this chapter such characterization is misleading if not wrong, for it is the relative length between V and coda rather than the length of the entire "diphthong" which distinguishes the two types of "diphthongs". The distinction, though considered contrastive only in the two pairs of "diphthongs" cited above in their scheme, is part of the language-specific regularities, and should be spelt out for the sake of either descriptive adequacy or proper pronunciation of Cantonese. Thus the "diphthongs" can be dichotomized according to the relative length of the first and second element:

- [13] LONG+SHORT: i:w, u:j, ε:w, ɔ:j, a:j, a:w
 SHORT+LONG: ej, ej, ow, ej, ew

There is, in fact, a not unreasonable way to characterize the kind of distinction in question in Daniel Jones' own terminology: i:w, etc. can be viewed as "falling diphthongs" (alias "descending/diminuendo diphthongs") and ej[:], etc., as "rising" diphthongs" (alias "ascending/crescendo diphthongs"), with the length distinction viewed as a realization of the more abstract distinction of "prominence". It is not un-ironical that while the analysis of rimes ending in -j and -w as diphthongs should be dismissed as missing the point that these phonetic diphthongs are phonologically speaking checked rimes, we find the terminology created for diphthong classification the most nearly appropriate for the characterization of the rime types in question.

Perhaps after all it is not calling them "diphthongs", but neglecting the paradigmatic relationship between -j and -w vis-à-vis syllable-final contoids that is undesirable. Recognizing the paradigmatic relation is not necessarily incompatible with the view that glide-checked rimes are diphthongs. With a little imagination, we may wonder why nasal-checked rimes, and indeed occluded rimes, cannot be viewed likewise. This is exactly the view expressed in Tung 1961 and 1964. Thus Tung (1964) writes:

Structurally, finals with [any final nasal] are comparable with those with [any (post-vocalic) non-syllabic vowel], thus may also be regarded as diphthongs. Considered together with the distribution of tones, [a final stop] is but allophonic to [the homorganic final

nasal].

However descriptively fruitful the diphthongal analogy is, we would not like to alter the widely accepted view of what diphthongs should be so as to justify a characterization of the rime types in question in diphthongal terms. Moreover, "rising/falling" and the two pairs of aliases do not show that the prominence contour is realized in the form of duration (rather than sonority, loudness, etc.). In a descriptive framework that recognizes morae, length distinctions result directly from moraic configurations. We would therefore like to characterize the rime types in question in moraic terms. Unlike the diphthongal characterization which by definition excludes the monophthongal unchecked vowel/rime, a moraic characterization of rime types will cover this third type of rime as well. Thus, to describe the rimes in [10a,b,c] I propose the following terminology, motivated by the moraic status of the constituent V and coda of the respective rimes:

[14]

MORAIC REPRESENTATION	CORRESPONDING VERBAL DESCRIPTION		
	Vowel	Coda	Rime
[10a]	Sesquimoraic	—	Uniform
[10b]	Full-moraic	Semimoraic	} Broken { Trochaic ¹ Iambic ¹
[10c]	Semimoraic	Full-moraic	

8.5 Mora versus feature

Until the present chapter length difference of the coda has been ignored. This is possible because coda length is non-distinctive on the one hand and has no obvious concomitant qualitative difference on the other. When we say that coda length is non-distinctive, what we mean is that it is predictable from the length status of the V. Since we have

¹Adapted from metrical terminology. 'Trochee' originally denotes the metrical foot -˘, i.e. a long syllable followed by a short syllable; 'iambus' originally denotes the metrical foot ˘-, i.e. a short syllable followed by a long syllable.

[15] Cd → [-α tense] / [α tense]—

[16] b. $\begin{array}{c} \text{V Cd} \\ | \\ \text{M M} \end{array} \rightarrow \begin{array}{c} \text{V Cd} \\ | \quad | \\ \text{M M} \end{array}$

 c. $\begin{array}{c} \text{V Cd} \\ | \backslash \\ \text{M M} \end{array} \rightarrow \begin{array}{c} \text{V Cd} \\ | \backslash | \\ \text{M M} \end{array}$

There are two possible ways of accounting for the V-length difference in [16] in moraic terms. One way is to regard the moraic configurations for the vowels in [16] as prime configurations, so that there is an intrinsic difference between the two vowels: whether or not the vowel is "ambi-moraic". Another way gains mileage from autosegmental phonological representations, viewing the input configurations in [16] as themselves deriving from simpler, more primitive configurations as in [17] and letting association take care of itself.

p.139

- [17] b. V Cd → [16b]
 M M
- c. V Cd → [16c]
 \
 M M

To ensure the subsequent association between V and the first mora, we need general principles such as that a mora must be associated with some segment and the cyclic application of inter-tier association, so that the first mora is not pre-empted by the onset.¹ The choice between these two treatments depends on how much cost one attributes to the higher-order principles. While I leave the choice open, we see that for either treatment the second mora is the key factor. The two vowels differ in whether the second mora is (also) associated to the V.

It is easy to see that whether a V is pre-linked to the second M is the moraic correlate of [*tense], at least in the case of checked rimes. It is only natural, therefore, to relegate the tense/lax distinction of V to moraic configurations, thereby reducing one distinctive feature and in turn withholding the distinction of four pairs of vowel qualities, namely e/ɛ, ɐ/æ, o/ɔ, ʊ/a, at the pre-moraic stage of organization.

Such relegation has the advantage of avoiding the inadequacy in the characterization of the difference between the two classes of vowels as [±tense] adopted tentatively so far. Note the pairing of the two classes of vowels:

- [18] **SHORT:** e ə ɒ ʊ
 LONG: ɛ: [ɛ̃:] ɔ: ʌ:

It involves qualitative as well as quantitative difference. While long/short is one of the conventional concomitant distinction of [*tense], the qualitative relation between the two classes is not the orthodox kind of relation: ε , α , \circ are not "executed with a greater deviation from

¹ As far as [17] is concerned, the convention proposed by some writers (Harakuchi 1977 and Clements & Ford 1979) that association be done from left to right is irrelevant: right-left association would result in the same configurations.

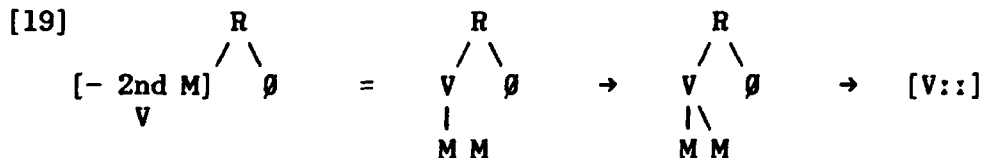
the neutral or rest position of the vocal tract" (SPE:324) than are e, ɛ, o. On the other hand, what is involved here is a regular difference in the single dimension of tongue height: a short V is always higher than its long counterpart. While nothing suggests that the raised position of short vowels follows from a tense/lax distinction, it is reasonable to regard such raising as resulting from brevity. Lingual onsets and codas prevail in Cantonese. Pronouncing a vowel most of the time involves a lower position of the tongue relative to adjacent sounds. While the tongue height (or "lowness" rather) of the tense vowels represents the target of tongue lowering, the lax vowels, owing to brevity, never reach the target. This kind of phenomenon occurs not infrequently in (short) glides in languages of the world. Cantonese itself provides good examples: a:j → [a:ǝ], ɔ:j → [ɔ:ǝ]. Compare also English "yard": [ǝ]/a:d/.

If the qualitative difference follows from a quantitative difference, the quantitative difference is arguably not primary either: it follows from the moraic representation. Thus, though [**long*] is pleasingly concrete, within a mode of description that includes the moraic representation, no matter whether [**long*] or [**tense*] is adopted, it has to be stipulated that one value of the feature triggers association with the second mora while the other value does not. It follows that any feature in terms of either length or tenseness is redundant, arbitrary and misleading, and in fact puts the cart before the horse.

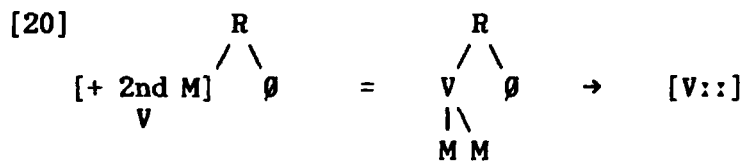
Just because we have justified the relegation of [*tense*]/[*long*] to moraic configurations, it does not follow that the two classes of vowels should not be differentiated in terms of a feature. On the contrary, the fact that whether a V is pre-linked to the second M is the moraic correlate of [**tense*]/[**long*] means that the moraic parameter involved is binary just as other mainstream features are. As such the binary choice can be formulated as [** 2nd M*], at least as a matter of notation. Such notation enables the moraic parameter to be fully incorporated into the motivated system of binary features adopted in this thesis for other areas in the phonology of Cantonese. For instance, a reformulation of the implicational relation "[*+high*] → [*+tense*]" in recognition of the moraic organisation is simply "[*+high*] → [*+ 2nd M*]".

An added advantage follows from a reformulation of [*tense*]/[*long*]

in moraic terms: we save one V-coda constraint out of four, namely LAX (*[-tense]Ø), which forbids unchecked short vowels. The reasoning works in the following manner. While [-tense] must mean shortness, [-2nd M] does not entail the latter. It so entails it only if the V is followed by a coda, when the second mora is associated with the coda and the coda only. [19] shows that something interesting happens when the V is unchecked.



Despite being [-2nd M], it results in the longest type of V. Admittedly [V::] is derivational-historically non-unique: [20] shows that the same [V::] results if the V is [+2nd M].



I argue that the non-uniqueness is harmless. The moraic organization of the language, requiring every syllable to be bi-moraic, predicts that the opposition [\pm 2nd M] be suspended for unchecked vowels. The fact that this prediction is borne out gives support to the moraic organization. In moraic terms the fact that [e, ɛ, o, ʊ] do not exist as a rime follows not from any constraint over the segmental configuration of the rime, but from the obligatory bi-moraic status of the wellformed syllable.

Along this line of thought, then, unchecked [e, ɛ, o, ʊ] will not be ruled out if mono-moraic syllables are somehow possible. As a matter of fact, tone bearing, mono-moraic "truncated" syllables can be demonstrated to occur in rapid casual speech as a result of contraction, resulting exactly in unchecked [e, ɛ, o, ʊ].¹ In a mora-free

¹ Details of the phenomenon will be given in Chapter 10.

framework of description the occurrence of these forms is counter-expectation and inexplicable. With the equipment of the moraic organization, on the other hand, both the non-occurrence of unchecked [e, ɛ, o, ɐ] in "full-form" syllables and their occurrence in contracted syllables fall naturally into place.

In the light of the discussion in this chapter, the table of rime at the end of Chapter Four needs two kinds of adjustment. First, the compatibility between [- 2nd M] and the highest degree of length, i.e. [V:], means that [- 2nd M] is qualitatively ambiguous. It follows that "e ɛ o ɐ" are over-specified, and therefore inadequate, symbols for the respective bundles of features that include [- 2nd M].¹ Second, as we have seen earlier, LAX is no longer needed. Apart from these two points, note that the length mark (:) is redundant, whether the analysis is moraic or not, and is all the more misleading in a mora-oriented analysis. While the length mark can be retained for practical reasons such as when doing linear transcription, I choose to lift it from my revised rime table [21]. The vowel qualities given only serve as reminders.

[21]	-θ	-j	-w	-m	-n	-ŋ
i	+	YOD	+	+	+	HI
u/y	+	+	LAB	LAB	+	HI
ε {	[+ 2nd M] (ε)	+	YOD	+	+	+
	[- 2nd M] (e/ε)	+	+	LAB	LAB	+
æ {	[+ 2nd M] (æ)	+	YOD	LAB	LAB	+
	[- 2nd M] (e/o/æ/ɔ) ²	+	+	+	+	+
ɔ {	[+ 2nd M] (ɔ)	+	+	LAB	LAB	+
a {	[+ 2nd M] (a)	+	+	+	+	+
	[- 2nd M] (ɐ/a)	+	+	+	+	+

¹ The lack of symbols for 'incompletely specified' segments is a sign of the superiority of features over atomistic segments as primes of contrast in the sound pattern of a language.

² Note that the 'short' counterparts of æ: and ɔ: are neutralized: [e] and [o] are in complementary distribution. Hence the present arrangement.

6.6 The place of morae in the syllable

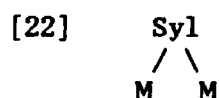
So far in this chapter we have made use of an autosegmental tier of morae to explain the following syllable-confined regularities:

- (1) Length complementarity between V and coda.
- (2) Rime isochrony.
- (3) Syllable isochrony.
- (4) Onset isochrony.
- (5) The 3:2:1 ratio of V length in different contexts.
- (6) The impossibility of unchecked [e e o ʊ] in full-form syllables.
- (7) The occurrence of unchecked [e e o ʊ] in contracted syllables.

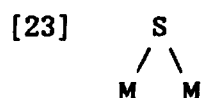
At this stage I assume that the moraic tier has established itself. The next question is, what is the position of the moraic tier in the representation of the syllable. In particular, what is its relationship to the hierarchical structure of the syllable?

We have been using the abbreviation "S" somewhat ambiguously, referring sometimes to the syllable as a whole and other times to only the segmental component of the syllable. To facilitate exposition we from now on use "Syl" for the inclusive syllable and reserve "S" for the latter reference.

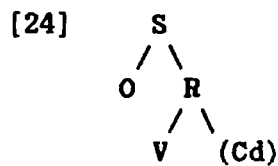
The first principle of the moraic organization of the syllable is that each syllable is bi-moraic. Thus the moraic formula of the syllable, in terms of autosegmental representation, is in the form of [22].



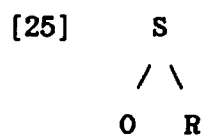
Since the segmental component of a syllable exhausts the temporal dimension of the latter, [22] implies [23].



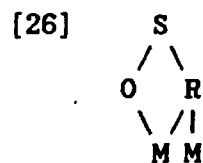
As far as S is concerned, tone and [*occl] are not included. The structure of S, therefore, looks as follows:



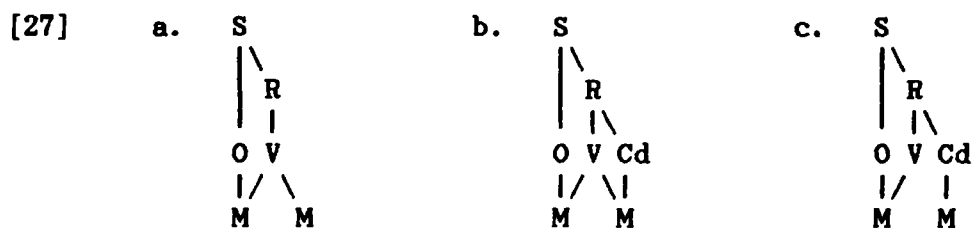
Consider [25], which is the top part of [24]:



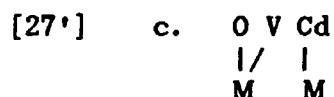
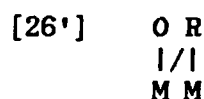
Given the resemblance between [23] and [25], we might want to ask how one is related to the other. I hasten to say that the resemblance is superficial. While [24] and therefore [25] represent a particular kind of tree diagram, namely one that signifies constituent analysis (Stewart 1976), [23] is not a tree diagram at all. Thus, M is not a constituent of S in the way that O and R are. Unlike the lines in [25], which signify constituency relationships, the lines in [23] signify "association" between segments and morae, which in turn represents, by convention, how the segments are temporally implemented, i.e. their duration. [22] and [23] account, for example, for syllable isochrony. On the other hand, [26], the moraic formula of S with details down to the level of O and R, accounts for onset isochrony and rime isochrony as well as syllable isochrony.



Lower down the hierarchy, at the level of terminal segments, there are three possible moraic arrangements, as we know already:



Putting together two kinds of information, namely constituency relations and moraic organization, which employ the same graphic devices of nodes and lines, is surely confusing. Any information not in focus at a given moment is therefore preferably omitted. For example, when tauto-syllabicity of the segments is not in focus, [26] and [27c] can be simplified to [26'] and [27'c] respectively:



[26'] and [27'] represent the usual kind of moraic representation that we have dealt with and will be dealing with. Lumping two kinds of representation together is seldom necessary because the hierarchical structure of S and the moraic representation have rather different functions. The formula for S serves to specify the wellformed segment sequences of the syllable, whereas the moraic representation serves to describe and explain realizational regularities, along the time dimension, of syllables and sequences of syllables. Some of these regularities have been covered in this chapter. Other phenomena that lend themselves to a moraic account will be dealt with in Chapter 7 and especially Chapter 10.

CHAPTER 7: THE SYLLABLE

Having looked at the various constituents of the syllable one by one we now move on to look at the syllable as a whole. Here we encounter the problem of syllabic nasals and the question of the restrictions concerning the combination of the various constituents of the syllable. Only when these are solved shall we be able to tell what makes a wellformed Cantonese syllable.

7.1 Syllabic nasals

In Cantonese a syllable can be made up, segmentally, of [m] or [ŋ] alone, thus [m̩] and [ŋ̩].¹ Given the canonical sequence of segments in a Cantonese syllable O+V(+Cd), which is collectively referred to as "S" from the last chapter onwards, we have difficulty fitting syllabic nasals into the standard formula for S.

Nasals occur normally as onsets or codas. Onsets are always momentary. In any case they will not be longer than half a mora. Codas, on the other hand, may be semimoraic or full-moraic. Of these three normal statuses of the nasal, the full-moraic coda is phonetically the most similar to the syllabic (and therefore bi-moraic) nasal. Moreover, [hm̩] and [hŋ̩], i.e. [m̩m, ŋ̩ŋ], exist as phonetic forms.² The phonetic resemblance of [m̩] and [ŋ̩] to long or full-moraic codas and the possibility of their taking an initial [h] together suggest that [m̩] and [ŋ̩] are rimes without a vowel.³ However I would like to argue against the treatment of [m̩] and [ŋ̩] as vowel-less rimes.

¹ Though the non-existence of [ŋ̩] in Cantonese is quite categorical, Hockett (1955:60, 1958:100) more than once includes it as one of the syllabic nasals in Cantonese, and actually uses it, rather than [m̩] or [ŋ̩], to illustrate his idea of 'syllable juncture'. Cantonese belongs to the class of languages which he has heard and 'done a little analytical work on'. The inclusion of [ŋ̩] probably results from a false generalization to the natural class of nasals.

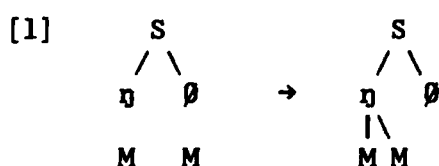
² Yuán et al 1960, Féng 1962, Hashimoto 1972 and Fung 1974 recognize [hŋ̩]. I observe that [hm̩] also exists. The two cases are similar and should be treated alike.

³ [m̩] and [ŋ̩] as rimes is the prevalent treatment on the Chinese mainland, probably owing to deference to Wong 1940 and Yuán et al 1960.

First, viewing [m̥] and [ŋ] as rimes with^{out} a V does not help fit them into the formula S=O+V(+Cd): the obligatory O and V are missing.

Second, there are reasons to believe that [hm̥] and [hŋ], unlike [m̥] and [ŋ], are not speech sounds. At any rate they are not the kind of syllable that can be used to build up lexical items. While [m̥] and [ŋ] have forms in T2, T4, T5 and T6 categorically,¹ the tone for [hŋ] is unclear. Fung (1974), Yuán et al (1960) and Hashimoto (1972) give only one tone [hŋ], suggesting that any difference in pitch in [hŋ] does not involve any lexical meaning. However, while Fung and Yuán et al take it to be in T6, Hashimoto takes it to be in T1. Féng(1962), on the other hand, includes both tones. The pitch shape of [hŋ], and also [hm̥], is in fact fairly variable. I argue that the different pitch shapes are the realization of intonation, not tone. Unlike [m̥] and [ŋ], [hm̥] and [hŋ] do not take any lexical tone. They are not proper syllables. As such they lie outside of the syllable phonology of Cantonese. Moreover, if [hm̥] and [hŋ] were accepted as the normal combination of onset and rime, the highly restricted onset-rime combination would have yet to be accounted for.

If [m̥] and [ŋ] are not vowel-less rimes, how should they be accounted for? We were first led into considering them as rimes on account of their phonetic resemblance to full-moraic -m and -ŋ. However, considered in moraic terms, m- and ŋ- are as likely as -m and -ŋ to be related to or responsible for, [m̥] and [ŋ]. Consider [1]:



Given the canonical bi-moraic status of the syllable, the association of both M's to η-, which is the only segment in the syllable, is automatic. The real problem we face is that this treatment again violates the formula S=O+V(+Cd): V is still missing.

¹ [m̥] is usually taken to occur with T4 only. However, because of the variation [m̥] ~ [ŋ] for what are regarded as prescriptively or underlyingly [ŋ], whatever tones co-occur with [ŋ] must also co-occur with [m̥]. The variation in question will be dealt with in Chapter 9.

To resolve this difficulty Chao (1947:22) posits underlying forms *mu:* and *ɲu:* for the two sounds in question:

After the initials *m* and *ɲg*, the function of *u* lies simply in the vocalization of preceding consonants, so that the whole syllable is pronounced as a syllabic nasal, thus *mu* [m̩], *ɲu* [ɲ̩].

Hashimoto (1972:173-4) dismisses Chao's treatment for the following reasons:

[W]ithin the synchronic description of Cantonese, the underlying forms /mu/ for [m̩] and /ɲu/ for [ɲ̩] do not seem to be sufficiently motivated. In addition, the analysis seems to add complexity to the overall description of the system. For example, we find in Cantonese a general tendency that single diffuse vowels do not occur after voiced initials. In fact, after the voiced initials, [m], [n], [ɲ], [l], only the single vowels [A:], [ɛ:] or [ɔ:] occur. Consequently the feature diffuse need not be marked with respect to these vowels when they occur as finals after the voiced initials. The following redundancy statement will predict it:

S24. [-cons] → [-diff] / [+cons]_{+voic} _____ +

But if [m̩] and [ɲ̩] are to be derived from /mu/ and /ɲu/, respectively, this general prediction will no longer hold.

But *mi:*⁶ occurs as the lexicalized form of the Cantonese equivalent of the musical note "me" in tonic sol-fa, and *ɲi:* appears in the onomatopoeic expression *ɲi:*⁴*ɲi:*¹*ɲɔ:*⁴*ɲɔ:*⁴ "murmur". So her S24 is not valid.¹ Besides, the synchronic motivation for Chao's treatment is two-fold. First, it eliminates the otherwise inexplicable segmental configuration of the syllables [m̩] and [ɲ̩]: mono-segmental and vowel-free. Second, if [mu:] and [ɲy:] (← *ɲu:*) do not occur, then the treatment will account for the phonetic non-occurrence of *m-* and *ɲ-* before *u:* but not before *i:* or *ɔ:*. However, this second motivation, and consequently Chao's analysis, works better for [ɲ̩] than for [m̩]. [ɲy:] is not at all a permissible sequence in the syllable. This results in the non-occurrence of [ɲy:] [ɲy:n] and [ɲy:t]. On the other hand, *mu:n* and *mu:t* occur in common lexical items. *mu:*, too, appears, in the

¹ *li:*¹, *ji:* (with various tones) and *wu:* (with various tones) also occur. Hashimoto (1972:162) explains one instance of *li:*¹ (her *ni:*) as alternating with *lej*¹ and another as an insignificant exception. As for *ji:* and *wu:*, they are *i:* and *u:* in her system. While alternation and exception are not convincing justifications for her S24 at all, I choose to appeal to *mi:* and *ɲi:* which she has overlooked.

loanword $\text{mu:}^1\text{fi:}^2$ "movie". mu: , therefore, is not a valid candidate as the underlying form for $[\text{ɱ}]$.

I propose to account for $[\text{ɱ}]$ in a different way from $[\text{ɳ}]$. $[\text{ɱ}]$ is to be treated as deriving from mi:m . As such it is at the same time related to an onset $[\text{m}]$ and a coda $[\text{m}]$. With the adoption of this treatment the second motivation mentioned above holds in an altered way: what is accounted for now is the non-occurrence of mi:m , not mu: .

With the obligatorily bi-moraic status of the syllable, Chao's "vocalization of preceding onset" can be re-formulated in more easily understandable terms as vowel deletion rules:

$$[2] \quad \text{u:} \rightarrow \emptyset / \text{ɳ} ___$$

$$[3] \quad \begin{matrix} \text{V} \\ [+high] \end{matrix} \rightarrow \emptyset / \text{m_m}^1$$

Derivation of $[\text{ɳ}]$ then takes the course depicted in [1], while that of $[\text{ɱ}]$ is as follows:

$$[4] \quad \begin{array}{ccc} \text{mi:m} & & \\ |/\backslash| & \rightarrow [3] \rightarrow & \begin{array}{cc} \text{m} & \text{m} \\ |/\backslash| & \end{array} \rightarrow \begin{array}{cc} \text{m} & \text{m} \\ | & | \\ \text{M} & \text{M} \end{array} = [\text{ɱ}] \\ \text{M} & \text{M} & \end{array}$$

7.2 Syllable-constituent combination restrictions

By analyzing $[\text{ɱ}]$ and $[\text{ɳ}]$ as underlyingly mi:m and ɳu: , we have preserved the following formation rules:

$$[5] \quad \text{Syl} \rightarrow \text{O} + \text{R} + \text{T} + [\text{*occl}] \\ \text{R} \rightarrow \text{V} (+ \text{Cd})$$

By our recognizing a sequential component S, the syllable can be characterized from another point of view:

$$[6] \quad \text{Syl} = \left[\begin{array}{c} [\text{*occl}] \\ \text{T} \\ \text{S} \end{array} \right]$$

¹ u:m is ruled out by the constraint LAB. See Section 4.3.

$$S = O + V (+Cd)$$

In [6] we represent the combination of [*occl], T and S, which are simultaneously executed, and that of segments, which are sequentially implemented, in two different graphic devices, assuming, by convention, that the horizontal dimension represents the flow of time. In Saussurean terms, O, V, and Cd are "syntagmatically" related. The relationship between [*occl], T and S, on the other hand, cannot be captured by Saussure's terminology. They are not in paradigmatic/associative relation. Nor are they in syntagmatic relation. This illustrates the validity of Jakobson's revision of the Saussurean dichotomy of paradigm vs syntagm to a hierarchical division as follows:

[7]

Relation between linguistic units	{	Combinational	{	Simultaneous
		Paradigmatic		Successive

In this perspective, O+R and V+Cd are combinations; so are T+[Occl]+S. And ignoring the hierarchy involving R, the syllable is the combination of T, [occl], O, V and optionally Cd, all of them "constituents" of the syllable (with the "sequential" connotation of the word "constituent" dismissed). Each of T, [occl], O, V and Cd is a system of paradigmatically related entities. As such each is a "paradigm" of a certain number of terms. A term in a paradigm combines with terms of other paradigms, resulting in some particular syllable. It is in this sense that we speak of the combination of syllable-constituents.

While the constituents combine relatively freely, they do not do so without restriction. For instance we have seen that there are strict constraints governing the combination of V and Cd. This section (7.2) looks at the restriction in the combination other than between V and Cd. For the sake of convenience of exposition the various kinds of (possible) restriction are divided into two groups, one involving [*occl] and the other, onsets.

7.2.1 [*occlusion]

Occlusion, i.e. [+occl], has restricted combination with tones on the one hand and with codas on the other. Such restrictions are dealt with in this section.

7.2.1.1 With tones

The relationship between tones and [+occl] is intricate. We have seen that the analysis of tones with occlusion as separate tones themselves, unrelated to their non-occlusive counterparts, is untenable. One reason is that exactly how many (plain) tones have an occlusive counterpart is not a question that has a straightforward and stable answer: it is awkward to have a paradigm "tone" having an undecided number of terms (9, 10, 11?), but it is quite all right to have unstable gaps existing in the combination of one paradigm with another, i.e. tone with [+occl]. However, even when [+occl] is now treated as a separate paradigm, we still need an adequate demarcation between systematic and accidental gaps, between illformed combinations of tones and [+occl] and wellformed ones that might or might not be regularly occurring. Linguists either ignore such demarcation or have simplistic ideas about it.

We can better understand the present state of affairs by looking at the historical derivation of tones and occlusion. Recall the Middle Chinese (MC) category of shēngs introduced in Section 3.1.1:

[8]	píng	shǎng	qù	rù
TRANSLATION	even	ascend	depart	enter
SUBSTITUTE LABEL	I	II	III	IV

At some stage of the development of MC, IV was realized as what I call "occlusion", i.e. the glottalization of final nasals, resulting in sounds which were no longer nasals but simultaneous oral and glottal stops:

[9]	Plain:	I, II, III
	Occlusive:	IV

In the intermediate systems between MC and certain modern dialects (including Cantonese), the pitch shape of III and IV is believed to have

been the same.¹ What distinguished IV from III was the presence of occlusion and the resultant shorter duration of the pitch-carrying portion of the syllable:

[10]	PITCH SHAPE		
	Plain	I II III	
	Occlusive	IV	

A later stage, intermediate between MC and present-day Cantonese (PCan), saw the "phonologization"² of the originally phonetic (i.e. non-language-specific) difference of high vs low pitch in the context of voiceless vs voiced onsets,³ resulting in the emergence of two "registers".

[11]	CONTEXT	REGISTER	[*occl]	PITCH SHAPE
	[-VOIC] onset	High	Plain	I II III
			Occlusive	IV
	[+VOIC] onset	Low	Plain	I II III
			Occlusive	IV

Rule [12] then destroyed the [*VOIC] difference among obstruent onsets:

[12] ONSET DEVOICING: [-SON] → [-VOIC] (Chen 1984:175)

Following this large scale reduction of the opposition [*VOIC] among

¹ Cf LY 1954.

² Hyman (1957): 'It will be generally assumed that the inventory of phonological features is identical to the inventory of phonetic features, and that languages implement these universal phonetic features in various linguistic ways. In other words phonetic features can be 'phonologized''(p.57-8) 'This process of phonologization, whereby a phonetic process becomes phonological,...'(p.171) Note the difference between this use of 'phonologize' and the use of the term to mean 'become phonemic', no longer allophones' (Lass 1984), for which sense I use the term 'phonemicize'.

³ The system of distinctive features to be used in this diachronic account represents a fundamentally different interplay of contrasts from that used for PCant. To avoid confusion I use italic capitals for the diachronically relevant features. For example, [*VOIC] ≠ [*voic].

onsets, there came the "phonemicization" of the register difference:

[13]

REGISTER	[*occl]	PITCH SHAPE
High	Plain	I _H II _H III _H
	Occlusive	IV _H
Low	Plain	I _L II _L III _L
	Occlusive	IV _L

As "register" is part and parcel of "pitch shapes", one should speak of six pitch shapes rather than three pitch shapes multiplied by two registers, thus:

[14]

	PITCH SHAPE					
Plain	I _H	II _H	III _H	I _L	II _L	III _L
Occlusive			IV _H			IH _L

Then the process took place in which IV_H took the pitch shape of I_H when the V is short:

[15]

	PITCH SHAPE					
Plain	I _H	II _H	III _H	I _L	II _L	III _L
Occlusive	IV _H [V]		IV _H [V:]			IH _L

[15] is the point of inception of modern descriptions of Cantonese, and the various labels translate into my system as follows:

[16]

	TONE SHAPE					
Plain	T1	T2	T3	T4	T5	T6
Occlusive	T1' [V]		T3' [V:]			T6'

The supposed complementary distribution of T1' and T3' is,

however, demonstrably defunct, because the T1' vs T3' distinction (that is, even if occlusive tones are recognized) has clearly phonemicized. Kao (1972) lists 22 T1' items with long V and five T3' items with short V. Yet the same writer asserts that "[T1'] occurs chiefly with syllables comprising a short vowel; [T3'] with those having a long vowel." I maintain that the 27 "deviations" from the said correlation are sufficient to disqualify it from being of any phonological significance: it has, at best, only statistical interest.¹

The defunct regularity seems to have misled Chao (1947:245), who generalizes that if the occluded syllable has an upper-tone and a short V, it is to be classified with T1; and if the occluded syllable has an upper-tone and a long V, it is in T3. He goes so far as to distort the facts and upset another regularity of the language in order to maintain the implication "T1' → short V". According to him, bi:t¹ ("certainly") has a short V.²(p. 21) Hashimoto (1972:177-8), who is in general very fond of this kind of diachronically motivated pseudo-regularity, also says that "[i]f all the exceptions are marked, the majority of [T1'] and [T3'] syllables need not be specified as high or mid [tone] in the lexicon, but only as non-low, just to be distinguished from [T6']." She accordingly formulates a redundancy statement to predict T1' or T3' in terms of tenseness of V.

Phonemicization of the T1' vs T3' difference results in [17]:

[17] T1(') T2 T3(') T4 T5 T6(')

This is the kind of distribution between tone and [occl] most widely accepted, and is the one represented in our RD. Wide acceptance, however, is no guarantee of validity. The derivation from [10] to [17] shows that the number of "occlusive tones", which are collectively the reflex of MC (shēng) IV, has been increasing. The development from [16]

¹ Even statistical interest is doubtful in the case of T1': according to Kao's own statistical figures, there are 69 regular T1' items (i.e. with short V) and 23 (not 22 this time) irregular T1' items (i.e. with long V). (p.160)

² [+high] vowels have no short counterparts. Even if [e] is taken to be the short counterpart of [i:], it occurs before -ŋ/k and -j only, not -t (except by virtue of the variation ek^et). Though the characterization of the V in bi:t¹ 'must' as short is clearly wrong, it is taken for granted in Yuán et al 1960:188, Chou 1968:12 and S Cheung 1972:18.

to [17] is of particular importance, because from [17] onwards occlusion is no longer intrinsically related to the tone shapes of T3 and T6, which are the reflex of MC III. If occlusion can extend from T3 and T6 to T1, there is no reason why other tones cannot be occlusive. Especially, following the extraction of [\pm occl] from the system of tones, we see clearly the existence of gaps for the combination of T2, T4, T5 with [+occl]. Some of these gaps, as we shall see, are beginning to be filled.

I have argued in Section 3.2.1 that T2 Switch has resulted in the lexicalization of some of its outputs. The lexicalized outputs of T2 Switch include occluded as well as unoccluded syllables. If loanwords and multi-syllabic words are included, lexicalized T2' items run into dozens. The most obvious indigenous, mono-syllabic examples include $\eta a:k^2$ "bracelet", $ti:p^2$ "card with notes", jok^2 "jade" and $m\ddot{o}:k^2$ "membrane". The fact that these items have graphically related non-T2 items either as their etymons, e.g. $\eta a:k^{3\rightarrow 2}$, or as their synchronic alternants, e.g. $ti:p^{3\sim 2}$, does not prevent speakers from storing the particular items in their brain as having T2. T2', i.e. T2 with occlusion, then, must be deemed a permissible combination. Thus:

[18] T1(') T2(') T3(') T4 T5 T6(')

The occurrence of T4' has been reported in Fàn 1979. It occurs as a result of mapping the tone melody T4+T2/1 (suggesting concrete, smallish objects in babytalk) onto a reduplicated monosyllabic noun, irrespective of its original tone. His examples are:

[19] $dz\ddot{a}:k^4 dz\ddot{a}:k^{2/1}$ "bird"
 $jok^4 jok^{2/1}$ "meat"

I observe the occurrence of T4' in onomatopoeic expressions. First, it occurs as a result of mapping the tone melody T4+T2 onto a reduplicated onomatopoeic item suggesting rhythmic sounds in the environment / $s\epsilon:\eta^1$ "sound":

[20] $\left. \begin{array}{l} b\left\{ \begin{smallmatrix} \text{e} \\ \text{o} \end{smallmatrix} \right\} p^4 b\left\{ \begin{smallmatrix} \text{e} \\ \text{o} \end{smallmatrix} \right\} p^2 \\ la:k^4 la:k^2 \\ g\ddot{o}:k^4 g\ddot{o}:k^2 \end{array} \right\} s\epsilon:\eta^1, \text{ suggesting, e.g., } \left\{ \begin{array}{l} \text{heartbeats} \\ \text{the cracking of bamboo} \\ \text{loud footsteps} \end{array} \right.$

Second, there is a class of onomatopoeic expressions adhering to the tone melody T4+T1+T4+T4:

- [21] fi:⁴li:l¹fɛ:t⁴lɛ:t⁴ (suggesting, e.g. sobbing)
 pek⁴lek¹pa:k⁴la:k⁴ (suggesting, e.g. slaps on the face)
 gi:⁴gi:l¹get⁴get⁴ "grumble"

Third, T4 is often used to represent low-pitch sounds in onomatopoeic expressions. An example is dek⁴da:k⁴, representing, for example, the ticks of the clock.

Then, in the interlanguage of speakers of Cantonese learning English, i.e. in Cantonese-English, all syllables after the last stress in an utterance are re-interpreted to bear T4. This also gives rise to T4':

- [22] T4 T4 T4
 phonetic, market, gossip.

If the examples in [19] [20] [21] are regarded as marginal and/or too few in number to establish T4', the scarcity of T4' in the core lexicon must be attributable to historico-accidental gaps in view of its pronounceability (illustrated in [22] as well as in the other examples) and its occurrence in less central parts of the lexicon. Hence the following tone-occlusion combination pattern:

- [23] T1(') T2(') T3(') T4(') T5 T6(')

Now the only tone that remains incompatible with occlusion is T5. Unlike T2' and T4', no occurrence of T5' of whatever status or by whatever process has been reported in the literature or observed by me. It is desirable, therefore, to regard the absolute non-occurrence of T5' as resulting from a genuine constraint against the combination of T5 and occlusion:

- [24] *T5'

The discovery of T2' and T4', or the weaker claim that their

scarcity represents accidental rather than systematic gaps, that T2' and T4' are not illformed, differentiates this thesis from all other works on Cantonese phonology. Though T2' and T4' have been inspired by the analysis that recognizes only six basic tones, with [\pm occl] extracted away from the tone system, nevertheless the occlusive-tone oriented analysis is by no means incompatible with the acceptance of T2' and T4'. One could treat T2' and T4' as additional tones, along with T1', T3' and T6'. These five occlusive tones, together with the six plain tones, would add up to eleven tones. Yet, as we all know, "nine" has been the stock number of tones for occlusive-tone oriented analyses. Perhaps the recognition of T2' and T4' might help persuade the occlusive-tone oriented analysts to rethink their position.

7.2.1.2 With rime

Unchecked and glide-checked rimes do not have the [\pm occl] distinction: glottalization applies only to final nasals. To describe the situation, we can say either that unchecked and glide-checked rimes are not permitted to combine with [\pm occl], or that the opposition [\pm occl] is "neutralized" when R = V or when Cd is [\pm cont]. However, given constraint [24], i.e. *T5', the fact that T5 does occur with unchecked and glide-checked rimes suggests that [-occl] is present rather than that the opposition [\pm occl] is neutralized. We should therefore need the following constraint:

$$[25] \quad * \left[\begin{array}{c} [+occl] \\ R \\ / \quad \backslash \\ V \quad \{ [+cont] \} \\ \quad \quad \emptyset \end{array} \right]_{Syl}$$

[25] looks a little clumsy. The clumsiness is partly due to the lack of interface between [\pm occl] and Cd in the structure of the syllable. In the wake of this consideration, and since [\pm occl] interacts with the coda-feature [\pm cont], an alternative interpretation of [\pm occl] suggests itself: it can be treated as a coda-feature too, when it would work with the same effect as [-/+ nasal]. This alternative has two advantages.

First, when R = V, the lack of coda automatically renders the

coda-feature [\ast occl] irrelevant. It follows that there is no longer the occasion to constrain against the combination of unchecked rimes with [+occl].

Second, since both [occl] and [cont] are coda-features, the interface problem no longer exists. An intra-paradigm implicational rule, just like the implicational rules for the paradigms tone and vowel, stated in terms solely of features, can serve to capture the incompatibility between [+cont] and [+occl], thus [+cont] \rightarrow [-occl]. In contrast, [\ast occl] as an IC of the syllable has no direct contact with the coda. The constraint has to make reference to non-sisters. In particular, the existence of a constraint between Cd (an IC of R) and [\ast occl], an entity beyond R) seems to endanger the status of R.

I shall first argue that the last advantage is not a genuine advantage. First, a constraint between Cd and an entity beyond R does not by itself endanger the status of R. All constituents of the syllable are inter-related. Hierarchical relations are justified by relative relatedness rather than by the presence or absence of relation.¹

Second, treating [\ast occl] as a coda-feature is not enough to avoid constraints between Cd and entities beyond R: as we shall see, an onset-coda constraint is needed anyway.

Third, if [occl] as a coda-feature avoids the interface problem for the [occl]-coda constraint, it creates an interface problem for the [occl]-tone constraint in [24], i.e. \ast T5', for [occl] will then no longer be a sister of tone.

Thus, as far as the interface problem is concerned, the two interpretations of [occl] are equally adequate and equally costly. The merit of the coda-feature interpretation, then, relies on the first

¹ Fudge (1986) even holds that 'constraints between Onset and Coda are irrelevant to the status of 'Rhyme''. Nevertheless I consider his claim too strong. In general constraints between non-sisters are less likely and less desirable than those between sisters. It is relative relatedness which ultimately determines hierarchicality. In regard to constraints, for instance, as we shall see, V-Cd constraints outnumber O-Cd and [Occl]-Cd constraints, and there exist O-R constraints, referring to R as a whole.

advantage. [25] has to refer to two kinds of situations: glide-checked rimes and unchecked rimes. This, however, stems from regarding the unchecked rime as consisting of nothing but a V. It follows from this treatment that Cd is an optional constituent of R, i.e. $R = V(+Cd)$. The disjunction $\{ \begin{smallmatrix} \emptyset \\ [+cont] \end{smallmatrix} \}$ in [24], however, suggests that \emptyset -coda, or zero coda, falls into a class with -j and -w. By representing zero-coda as $[+cont, -cor, -lab]$, $[+cont]$ defines a class that includes the zero-coda as well as -j and -w. The constraint in question can then be represented simply as [26]:

$$[26] \quad * \begin{bmatrix} [+occl] \\ [Cd \\ [+cont]] \end{bmatrix}$$

As a result of this re-interpretation of the coda-less rime, even the first advantage of the coda-feature interpretation is no longer an advantage. The two alternatives are equally adequate and equally costly. I choose to continue treating $[occl]$ as an IC of the syllable, mainly for the pragmatic reason of setting a link between the occlusive-tone oriented analysis, where $[occl]$ would be a tone-feature, and occlusive-coda analysis, where $[occl]$ would be a coda-feature.

This reinterpretation also implies that Cd is no longer optional: it is now obligatory. The obligatoriness in turn bears on the syllable formation rules. There will be further discussion of this topic in Section 7.3 below.

7.2.2 Onset

Restrictions on the combination of onset with tone, coda and rime have been reported in the literature. Not all of these restrictions are valid. We shall deal with these three kinds of restriction one by one.

7.2.2.1 With tone

7.2.2.1.1 Nasals and l-

Hashimoto (1972:146-7) posits the following "redundancy statement":

[27] [+syll] → [lo] / [^{+cons}_{+voic}]

which, in her system, means that a syllable having m-, n-, l- or ŋ- must be in T4, T5 or T6, i.e. one of the lower tones. She herself, however, lists some fifty items that violate [27]. And the list is by no means exhaustive, as she is well aware, judging from her use of such expressions as "include", "such" and "etc.". Her classification of these items into exceptions/phonologically colloquial, colloquial morphemes, particles, onomatopoeic words and loanwords does not seem to me to have served to rescue [27]. Her way of handling the fifty counter-[27] items is to restrict the applicability of [27] to "literary morphemes". But since there is no independent and rigorous criterion of "literariness", the argument is circular.

The difficulty stems from her intention to represent diachronic regularities in the synchronic system. It is now commonplace to note that while diachronic regularities often provide clues for the extraction of synchronic regularities and vice versa, it is methodologically wrong to confuse the two. In the present case, the sound changes involved have to do with [11] [12] and [13]. The combined effect of [11] and [12] is that at the stage immediately before [13], sonorant onsets did not co-occur with the high-register pitch shapes, thus:

[28]	ONSET CLASS	REGISTER
	[-SON]= [-VOIC]	high/low
	[+SON]= [+VOIC]	low

This is so because even prior to that stage, [+SON] had implied [+VOIC]: unlike the [-SON] onsets, whose former [*VOIC] distinction now manifested itself as low or high register respectively, [+SON] onsets had had no voiceless counterparts which would now be in the high register. Following the phonemicization of the high/low register difference, resulting in [13], or less misleadingly [14], we expect (i) new items to emerge to fill the historico-accidental gaps, and (ii) some reflexes of the sonorant items of the earlier stage to appear in the higher tones contrary to the diachronic regularity. Both of these expectations are now demonstrably borne out by the facts of present-day Cantonese. It

[+VOI, -SON] onset (or low register), as depicted in [30]:

[30]	$\begin{bmatrix} 0 \\ -SON \\ +VOI \end{bmatrix}$	I/II	III/IV
		[+ASP]	[-ASP]

[30] translates into the post-[13] system as [31] and in turn into the present-day system as [32]:

[31]	$\begin{bmatrix} 0 \\ -SON \end{bmatrix}$	I _L /II _L	III _L /IV _L
		[+ASP]	[-ASP]

[32]	$\begin{bmatrix} 0 \\ -cont \end{bmatrix}$	T4/5	T6
		[-voic]	[+voic]

There are those who believe in the present-day validity of [32]. For instance Hashimoto (1972:147) writes:

Syllables with (...) unaspirated stop or affricate initials do not occur in [T4] or [T5]; while those with (...) aspirated stop or affricate initials do not occur in [T6].

She formulates "redundancy statements" to provide for this alleged present-day regularity. Kao (1971:126) holds a similar view. Both Hashimoto and Kao, however, are aware of the existence of exceptions. Hashimoto lists (non-exhaustively) six examples of items with a [+voic, -cont]¹ onset and T4, and four items with a [-voic, -cont] onset and T6'. The combination of [-voic, -cont] onsets with T6 is in fact not confined to the occluded environment. Consider the following loanwords taken from S Cheung 1972 [33a] and Kiu 1977 [33b]:

[33]	a.	kum ⁶ mi:l ¹ sen ²	← commission
		ko:n ⁶ de:n ¹ sa:2	← condenser

¹ My features.

b.	pow ⁶ tɛ:n ¹ sow ²	← potential
	kɐm ⁶ pi:w ¹ ta:2	← computer
	pi: ⁶ di: ⁶	← paediatrics
	ji: ⁶ kɔ:n ⁶	← economics

The last example appears also in Y Cheung 1986:48. These loanword forms, in turn, can be demonstrated to have derived from Cantonese-English interlanguage forms resulting from reinterpreting the neutralized English intonation tone-wise such that all syllables before the first stress in an utterance are rendered in T6¹. By virtue of this general rule, any onsets with T6 are pronounceable and can appear in loanwords potentially. In addition to [33], I also observe the following loanwards:

[34]	ti:w ⁶ tɔ:l	← tutorial
	pow ⁶ li:t ¹	← political science
	kow ⁶ [ʔ]ɛ:t ¹	← co-educational
	pɛ: ⁶ sɛ:n ¹	← percent

With the existence of all these examples, including those furnished by Hashimoto herself, the alleged regularity as shown in [32] no longer holds. There is, however, complete absence of the combination of [-cont, +voic] onsets with T5. Thus, rather than dismiss [32] in its entirety, we qualify its scope, resulting in the following constraint:

$$[35] \quad * \begin{bmatrix} \text{T5} \\ \begin{bmatrix} \text{-cont} \\ \text{+voic} \end{bmatrix} 0 \end{bmatrix}$$

7.2.2.2 With coda

Kao posits the following constraint on the combination of onset and coda:

¹ A similar rule has been formulated by Luke (1984:194-5) at the level of word (as opposed to utterance).

[36]		-m/p	-w
	b- p- m- f-	-	+
	gw- kw- w-	-	-

She thinks that "[t]he only exception to this rule is the loanword [bem] from English *pump*", but in fact the following common expressions exist:

[37]	beŋ ¹ be m lbo:1	"ping-pong"
	m l m l	"food" (babytalk)
	bep ⁴ bep ² sɛ:ŋ ¹	"sounds resembling heart-beats"
	wow ¹	"bark"
	bi:p ¹	"beep"

mim, the underlying form for [m] adopted in this thesis, also goes against the constraint. All these examples call for the revision of [36] in favour of [38].

[38]		-m/p	-w
	b- p- m- f-	+	+
	gw- kw- w-	-	+

The non-occurrence of f- with -m/p (and the occurrence of few in very common items such as in T1/2/4/6 and the less common fow in tow¹fow² "TOEFL" = "Test of English as a Foreign Language") corroborates our earlier classification of f- not with b- p- m- but with gw- kw- w-:

[39]		-m/p	-w
	b- p- m-	+	+
	gw- kw- w- f-	-	+

Table [39], however, is misleading in that it does not reflect the different distribution of gw- and kw- vis-à-vis w- and f-: co-occurrence of gw- or kw- with any labial coda is not permitted. Hence the revision of [39] as [40]:

[40]		-m/p	-w
	B (b- p- m-)	+	+
	F (w- f-)	-	+
	Q (gw- kw-)	-	-

To characterize the restriction at work, I formulate the following constraint:

$$\begin{array}{c}
 [41] \text{ LABIODENTALS Constraint: } *S \\
 \begin{array}{c}
 \begin{array}{c} \diagup \\ R \\ \diagdown \end{array} \\
 \begin{array}{c} \diagup \\ X \\ \diagdown \end{array}
 \end{array}
 \begin{array}{c}
 \begin{bmatrix} -cor \\ +ant \\ -dist \\ \langle +cont \rangle \end{bmatrix}
 \begin{bmatrix} +lab \\ \langle -cont \rangle \end{bmatrix}
 \end{array}
 \end{array}$$

7.2.2.3 With rime

How freely onset combines with rime is an important and interesting question. The question, however, could be misleading owing to the inherent ambiguity of the expression "freedom of combination". All those works on Cantonese phonology which provide an allegedly exhaustive list of "occurring syllables" are tacitly committed to drawing a sharp line between the occurring and non-occurring syllable. For the authors of these works, free/missible vs unmissible combinations are synonymous with occurring vs non-occurring combinations. However, non-occurrence of a syllable merely means the non-exploitation (which could be temporary) of the syllable in the lexicon: it tells nothing about the potentiality of such exploitation. The lexicon is the most volatile component of a language and the least rule-governed. Words, and therefore syllables, enter and leave the lexicon. Once compiled, the "exhaustive" list of occurring syllables is quickly outdated. The technical difficulty, however, is not the most important reason why I take issue with the [\neq occurring] distinction. The heaviest blow to [\neq occurring] is the fact that it is no more than a historical accident: it is not related to linguistic competence. As such it is not a meaningful distinction as far as phonology is concerned. What is phonologically relevant and more meaningful is the distinction between wellformed/theoretically possible vs illformed/theoretically impossible syllables, or combinations of the constituents of the syllable. Closely related to this distinction is the distinction of systematic vs accidental gaps. Thus, a non-occurring combination may nevertheless be wellformed, the gap being accidental, not systematic.

The foregoing comments apply to any combination of syllable-constituents, but the confusion between [*occurring] and [*wellformed] is most serious in the case of the combination of O and R. Probably because of the large paradigm size of both O and R, and the high rate of under-exploitation of all the logically possible combinations, here the line between systematic and accidental gaps is most difficult to draw. I have anyhow made an effort to draw it. In order to make an adequate distinction of [*wellformed], I adopt the following strategies.

First, I depend on a maximally expanded inventory of occurring syllables drawing from the lexicon of myself as an observing native speaker of Hong Kong Cantonese. I have to take this course because none of the published syllabaries is "complete" from my viewpoint, and most of them are dated. Thus, I recognize as occurring some syllables that Zhāng 1983 and Bauer 1984, which provide the most nearly adequate syllabaries, do not recognize.¹ The importance of a more comprehensive syllabary cannot be over-emphasized, in view of the general principle that occurrence implies wellformedness.

Second, I draw not only on core lexical items but also on peripheral and marginal ones, such as loan words and onomatopoeic expressions. Despite the non-centrality of their position in the lexicon, I hold that occurrence of a syllable in such items also implies wellformedness.

Third, I make reference to the pronounceability of a syllable, drawing on the (early-stage) interlanguages spoken by native speakers of Cantonese learning other languages, basically Mandarin and English. The general principle is that pronounceability is correlated with wellformedness.

Fourth, I observe that except for the disparity between Q- (gw-, kw-) and F- (w-, f-), there is a strong tendency for homorganic onsets to have the same distribution with respect to rime. On the basis of this

¹ Zhāng's syllabary suffers from the non-recognition of ε:w, ε:m/p and ε:n/t as regular rimes (though she mentions in passing four syllables bearing some of these rimes) and from insensitivity to loanwords. Bauer 1984 suffers from the fact that he possesses a fairly limited lexicon. Moreover, he could have referred to Zhāng's work but he did not.

tendency, I assume that any discrepancy among homorganic onsets (Q- and F- are deemed heterorganic for this purpose) represents an accidental gap.

Fifth, I formulate constraints to account for systematic gaps. In view of the principle that the more general the constraint is and the more readily the constraint can be stated in terms of natural classes, the more likely is the existence of a systematic gap, the formulated constraints serve as a yardstick of the plausibility of systematic gaps.

With the adoption of these strategies, I arrive at the table of O-R combination [42]. An explanation of the notations follows.

[42]	Q	F	B	D	S	G
a:	+	+	+	+	+	+
a:j	+	+	+	+	+	+
a:w	Cd		+	+	+	+
a:m	Cd	Cd		+	+	+
a:n	+	+	+	+	+	+
a:ŋ	+	+	+	+	+	+
ɐj	+	+	+	+	+	+
ɐw	Cd	+	+	+	+	+
ɐm	Cd	Cd	+	+	+	+
ɐn	+	+	+	+	+	+
ɐŋ	+	+	+	+	+	+
ɛ:	+	+	+	+	+	+
ɛ:w	Cd	new	+	+	new	new
ɛ:m	Cd	Cd	new	+	+	+
ɛ:n	Fr	+	+	new/↓	new/↓	new/↓
ɛ:ŋ	Fr		+	+	+	+
ej		+	+	+	+	+
en	↓	↓	↓	↓	↓	↓
eŋ	+	+	+	+	+	+
œ:	Q~G	+	+	+	+	+
œ:n	Fr	↓	↓	↓	↓	↓
œ:ŋ	Fr			+	+	+
ej	[e]	[e]	[e]	+	+	+
en	[e]	[e]	[e]	+	+	[e]
ɔ:	+	+	+	+	+	+
ɔ:j	Q~G			+	+	+
ɔ:n	Q~G/↓	↓	↓	↓	↓	↓
ɔ:ŋ	+	+	+	+	+	+
ow	Cd	+	+	+	+	+
om	Cd	Cd	+	+	new	+
oŋ	Q~G	+	+	+	+	+
i:	+	+	+	+	+	+
i:w	Cd	+	+	+	+	+
i:m	Cd/Fr	Cd	+	+	+	+
i:n	Fr	+	+	+	+	+
y:/u:	+	+	+		+	
u:j	+	+	+	YOD	YOD	YOD
y:n/u:n	+	+	+	+	+	+

D = Dentals, i.e. d-, t-, l-.

S = Sibilants/palatals, i.e. dz-, ts-, s-, j-.

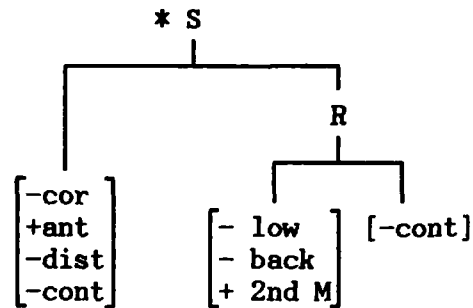
G = Gutturals, i.e. g-, k-, ŋ-, h-.

+ = Wellformed by virtue of occurrence.

Cd = Illformed by virtue of the O-Cd constraint LABIODENTALS, i.e. [41].

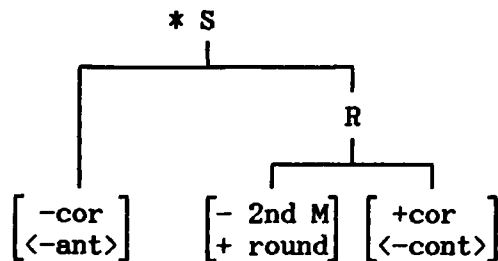
Fr = Illformed by virtue of constraint [43] against the combination of gw- , kw- with i:m, i:n, ε:n, ε:ŋ, œ:n, œ:ŋ.

[43] FRONT-V Constraint:



[e] = Illformed by virtue of constraint [44] against the combination of the onset groups Q, F, B with the rimes ej, en, and that of G with en.

[44] [e] Constraint:



YOD = Illformed by virtue of the V-Cd constraint YOD¹ against, among other things, [y:j], in association with the realization rule (to be formulated in Section 8.3) which provides that y:/u: is to be realized as y: after non-labials.

I hold that the four constraints above account for all the truly ill-formed O-R combinations. Among the remaining non-occurring O-R combinations, the following markings denote particular reasons why a certain combination is not to be regarded as illformed.

¹ See figure [23], Section 4.3.

"↓" concerns the rimes $\varepsilon:n$, en , $\alpha:n$ and $\upsilon:n$. By way of the variation $-ŋ \sim -n$, the n -checked syllables occur as variants of the $ŋ$ -checked syllables. The fact that the gaps are beginning to be filled in this way suggests that they are historico-accidental, not systematic.

"new" refers to newly recognized rimes, including $\varepsilon:w$, $\varepsilon:m$, $\varepsilon:n$ and om . These rimes are assumed to have emerged not long ago. After they established themselves as wellformed rimes, new lexical items have gradually emerged to fill the combination gaps between onsets and these new rimes. Thus we have reasons to believe that the existing gaps involving these rimes result from a time lag between the establishment of the rimes and their combination with the whole range of onsets or onset groups.

"Q~G" has to do with variation [45] to be elaborated in Chapter 9.

$$[45] \quad \begin{bmatrix} gw \\ kw \end{bmatrix} \sim \begin{bmatrix} g \\ k \end{bmatrix} / \quad \begin{matrix} v \\ [+round] \end{matrix}$$

There are signs that gw - and kw - are giving way to g - and k - before $\upsilon:$ and $u:$. We have reasons to believe that the language is developing towards a stage where a constraint in the form [46] is to operate:

$$[46] \quad * \begin{bmatrix} -cor \\ +ant \\ +dist \\ -cont \end{bmatrix} \begin{matrix} v \\ [+round] \end{matrix}$$

Owing to this emerging regularity, $Q\alpha:$, $Q\upsilon:j$ and Qon , though not exactly illformed at the present stage, nevertheless are not combinations favoured by the language. This explains why they have not been exploited in the lexicon.

The empty slots that remain are the truly "accidental" historico-accidental gaps in the sense that there is no particular explanation for their non-occurrence. I cite elementary-stage Cantonese-English (unmarked) and Cantonese-Mandarin interlanguage pronunciations to establish their pronounceability:

Fa:w	<i>flower, vowel.</i>
Ba:m	<i>balm, palm.</i>
Fa:ŋ	<i>thank, Wang, Wrangler.</i>
Qe:j	<i>equation. Mandarin: guèi "expensive".</i>
Fu:ŋ	<i>work.</i>
Bæ:ŋ	<i>iceberg.</i>
Fɔ:j	<i>voice</i>
Bɔ:j	<i>boy (marginal loanword), poison.</i>
Dy:	<i>Mandarin: lǜ "green", nǚ "female".</i>
Gy:	[No interlanguage examples. But gy: is u ⁵ ed in the imitation of an expression in the Shùndé dialect (of the Yuè dialect group) in the Cantonized way, thus ɲɔ:ŋ ⁶ gy:lgy:l "stupid".]

7.3 Characterizing the wellformed syllable

7.3.1 The syllable structure revisited

Following the conception of the lack of a coda as the presence of a zero coda characterized as [+cont, -cor, -lab], Coda becomes an obligatory constituent of the syllable; and following Jakobson's taxonomy of formal relationships as represented in [7] above (Section 7.2.0), the Cantonese syllable is the simultaneous combination of tone, [occl] and S, and in turn S is the successive combination of O and R, and in turn R, that of V and Cd. Following the convention of depicting simultaneity vertically and successiveness horizontally, the syllable in the inclusive sense, Syl, can be depicted as [47], and S as [48] or (with the hierarchicality ignored) as [49].

$$[47] \left[\begin{array}{c} T \\ [occl] \\ S \end{array} \right]_{Syl}$$

$$[48] \begin{array}{c} S \\ | \backslash \\ R \\ | \backslash \\ O \ V \ Cd \end{array}$$

[49]
$$\begin{array}{c} S \\ \triangle \\ O \quad V \quad Cd \end{array}$$

The difference between simultaneity and successiveness, however, is not only secondary to the paradigm vs combination dichotomy (as Jakobson argues), but is sometimes unclear or unnecessary. Take the SPE characterization of affricates as $\begin{bmatrix} - & \text{cont} \\ + & \text{del rel} \end{bmatrix}$. While a bundle of distinctive features in general represents simultaneous components of a segment (hence "feature columns" by convention), [-cont] and [+del rel] are not really simultaneously implemented. Compare a possible auto-segmental representation of affricates as [50], which spells out the successiveness of the stop and the release.

[50]
$$\begin{array}{cc} [-\text{cont}] & [+ \text{cont}] \\ & \backslash / \\ & C \end{array}$$

The simultaneity guise of the SPE representation is possible because the domain of every feature is a well-defined unit, namely the phoneme-size segment. Successiveness occurs within the confines of the segment and the exact locus of a particular feature is predictable and is therefore redundant information.

Now consider the Cantonese syllable. Both S and, to a lesser extent, T, have the entire syllable as their locus. Despite the localized realization of [occl], since there is only one such realization in each syllable, we have no difficulty viewing it as a syllable-level property, especially as it "succeeds" neither S nor T. On the other hand the successiveness of O and R is self-evident. This very successiveness has led us to treat the O-R combination in a different way from the simultaneous combination of T, [occl] and S. But a moment of thought reveals that the successiveness of O and R is predictable, just as that of [-cont] and [+del rel] is predictable in the SPE characterization of affricates. It follows that the successiveness of O and R need not be overtly represented. Thus, instead of the horizontal representation of S as O+R, one

may also envisage a vertical representation $\begin{bmatrix} O \\ R \end{bmatrix}_S$, despite its simultaneity guise. By the same token, $\begin{bmatrix} V \\ Cd \end{bmatrix}_R$ gives the same result as

"R=V+Cd". As the argument goes, both V and Cd are rime-level properties and both R and O are S-level properties. This in turn means that all of V, Cd and O (along with R) can be viewed as S-level properties. But S itself is a syllable-level property. So it follows that O, V and Cd are syllable-level properties too. In general, then, because of the predictability of the locus of realization, all constituents of the syllable are properties of the syllable. Thus:

$$[51] \quad \left[\begin{array}{c} T \\ [\text{occl}] \\ O \\ V \\ Cd \end{array} \right]_{\text{Syl}}$$

In [51], the internal hierarchy is not reflected. When the simultaneity vs succesiveness difference is not at issue, S is not really needed in the characterization of Syl. We come back to the following formation rules for Syl, with the simultaneity vs successiveness difference ignored:

$$[52] \quad \text{Syl} \rightarrow T + [\text{occl}] + O + R \\ R \rightarrow V + Cd$$

This set of horizontally written formation rules is isomorphic with the following columnar representation:

$$[53] \quad \left[\begin{array}{c} T \\ [\text{occl}] \\ O \\ \left[\begin{array}{c} V \\ Cd \end{array} \right]_R \end{array} \right]_{\text{Syl}}$$

This line of reasoning can be carried still further. Note that each of T, O, V and Cd is itself decomposed into binary features. Cd, for example, is a complex of [*cor, *cont, *lab]. It follows that the Cd in [53] can be detailed as [54]:

$$[54] \quad \left[\begin{array}{c} [*cor] \\ [*cont] \\ [*lab] \end{array} \right]_{Cd}$$

O, V and Cd can be expanded likewise, resulting in the ultimate representation of the syllable. To characterize the wellformed syllable, we need this representation of the syllable as a hierarchical complex, together with a set of combinational restrictions.¹

This non-linear approach to the syllable is characteristic of both indigenous Chinese phonology and Firthian prosodic analysis. Summarizing "Chinese phonological theory", Halliday (1981) writes:²

The phonology remained a phonology of the syllable, always analysed into initial and final, with the initials *classified* by place and manner of articulation and the finals by rhyme, vowel grade, labialisation and tone. There was a clear distinction between systematic and accidental gaps: between syllables regarded as theoretically impossible, and those regarded as theoretically possible but happening not to occur. (...) The way of looking at the syllable was akin to that of Firthian prosodic phonology (cf. Palmer 1970). There was no concept of further segmentation into phoneme-type units, but rather systems of contrasting features which interacted without being rigorously localized.

Thus, syllable-constituents (T, [occl], S, O, R, V and Cd) and distinctive features alike serve to *classify* the syllable, and any localization of these classificatory factors is derived, not primary.

7.3.2 Summarizing the wellformed syllable

- 1) The set of wellformed Cantonese syllables is generated by the combination (not necessarily successive) of the paradigms T, [occl], O, V and Cd, subject to various restrictions over the combination.
- 2) Each paradigm, except for [occl] (which is binary already), results from the various combinations of a fixed set of binary features.
- 3) To capture the close relationship between V and Cd, V+Cd is regarded as a higher-order paradigm referred to as R.³

¹ Taken in the widest sense, 'combinational restrictions' include the implicational relationships that hold among the binary features within the lowest-level paradigms T, O, V and Cd.

² The emphasis is mine.

³ Since the characterization of the wellformed syllable makes no reference to S as a whole, S has no special status as far as the wellformedness condition of the syllable is concerned. It is, however, a

- 4) The features for the respective lowest-level paradigms are as follows:

T: [high] [rising] [extreme]¹
 O: [cor] [ant] [dist] [cont] [voic]
 V: [high] [low] [back] [round] [2nd M]
 Cd: [cor] [lab] [cont]

- 5) According as the restrictions govern intra- or inter-paradigm combination, they are regarded in this thesis as "implicational relations" or "constraints".

- 6) The following implicational relations exist:

T: [+rising] → [+high]
 O: nil²
 V: [+low] → [-round]
 [+back] → [+round]
 [+high] → [+ 2nd M]³
 Cd: [+cor] → [-lab]

- 7) The following constraints exist:

[occl]-T: T5', i.e. $\left[\begin{array}{c} [-\text{extreme}] \\ [+rising] \\ [+occl] \end{array} \right] T$

[occl]-Cd: $\left[\begin{array}{c} [+occl] \\ [[+cont]]Cd \end{array} \right]$

V-Cd:⁴ YOD, LAB, HI

convenient collective term for the segments of a syllable, i.e. the syllable-constituents which are sequentially implemented and exhaust the temporal dimension of a syllable.

¹ [falling] would be needed to differentiate the variant forms of T1 and T4.

² However, [*dist] is neutralized when [+cor] or [-ant].

³ Apart from the implicational relations, [*back] is neutralized in the environment /

$$\left[\begin{array}{c} +\text{round} \\ \{-2\text{nd M}\} \\ +\text{high} \end{array} \right]$$

⁴ See figure [23], Section 4.3.

$$\text{O-T: } * \left[\begin{array}{c} \left[\begin{array}{c} [-\text{extreme}] \\ [+rising] \end{array} \right] \text{T} \\ \left[\begin{array}{c} [-\text{cont}] \\ [-\text{voic}] \end{array} \right] \text{O} \end{array} \right]$$

O-CD: LABIODENTALS, i.e. [41].

O-R: FRONT-V ([43]), [ə] ([44]).

CHAPTER 8: REALIZATION¹

This chapter deals with the phonetic realization of the more or less abstract phonological contrasts we have established so far, including the noted variability in such realization. "Realization" is ultimately the realization of abstract/formal distinctions, such as the kind represented by binary distinctive features. The realization of features, however, is often context-sensitive, and features pertaining to the same paradigm interact closely together to characterize and contrast terms within that paradigm. For ease of exposition, therefore, we shall be, most of the time, looking at the realization of the respective terms within each of the paradigms [±occl], T, V, Cd and O, which paradigms can be thought of as the lowest-level constituents of a syllable — rather than look at the respective features that cross-classify such terms. A particular kind of realizational regularity, namely lip-rounding harmony, is not "localized", and will be dealt with in a separate section. Much of the information given in this chapter has been alluded to in previous chapters, but the presentation here is more complete and better organized. As for the special cases of the realization of *mi:m* as [m̥] and *ɲu:* as [ɲ], since they have already been dealt with in Section 7.1, they will not be repeated in this chapter.

8.1 [Occlusion]

For this particular paradigm, the binary feature coincides with the paradigm, while the feature-values coincide with the terms in the paradigm. The effect of [occl] is four-fold. Thus,

- [1] [±occl] implies: (i) [± complete glottal closure at coda]
(ii) [± nasality of coda when [-cont]]
(iii) [± short pitched portion of syllable]

¹ The content of this chapter has benefited considerably from the five hourly sessions devoted to Cantonese as part of the practical course Phonological Analysis during the academic years 1984/85 and 1985/86. The course was taught by Dr. J.C. Wells; the participants were postgraduate linguistics/phonetics students of University College London and the School of Oriental and African Studies. Participants in the Cantonese sessions included two native speakers of Cantonese. Any error in description of course remains my own responsibility.

(iv) [± possibility of falling tone shape]

Kao (1972:61) remarks that "syllables with stop codas generally occur in level tones". In her description, "level" applies only to T3, T6 and the HE variant of T1, and so it follows from the above remark that only T3, T6 and HE are compatible with occlusion. However, it has been established in Section 3.3 that T4, like T1, has an even and a falling variant, and that T4 is also compatible with occlusion. As a matter of fact the LF variant of T4 is not compatible with occlusion. Thus a more general statement is that [+falling] implies [-occl], or [+occl] implies [-falling]. That this has to do with (iii) is obvious. One might therefore be tempted to generalize that [+occl] implies [-contour], suggesting that shortness of the pitch-carrying portion prevents a pitch contour from being realized as such. But T2, a sharp rise, readily co-occurs with occlusion, thus violating the last implication. Kao (1971:61) writes, "Only in their changed-tone forms will syllables with stop codas occur in any but the level tone." By characterizing the environment for the said violation as "changed-tone forms" she is tacitly justifying the regularity "[+occl] → [-contour]". It has been established in Section 3.2.1, however, that there exist lexicalized T2' items. It follows that an explanation of the said violation by appeal to "changed-tone forms" is not valid, and the rule [+occl] → [-contour], in this simple form, though phonetically motivated, would not work. The difference between T2 and T1/4 is that while the rise of T2 is an obligatory and the most salient property of T2, the fall of T1 and T4 is just one of the two possible realizations of the tone: they have an alternative even tone-shape at their disposal. Taking both phonetic motivation and lexical distinctness into consideration, we arrive at the following alternative formulation of implication (iv):

[2] (iv'): [+occl] → [-contour] when [*contour] is not contrastive

Rephrased as [2], we can see that the fourth effect of occlusion is motivated by its third effect: shortness does not favour pitch contour.

But (iii) itself follows from (i): that [+ complete glottal closure] implies the lack of vocal fold vibration is obvious. It follows that (i) implies (iii) and is indirectly responsible for (iv) at the same time. As for [± nasality of coda], it is often regarded as the realization of

[*occl] when the coda is [-cont], but this is to ignore the simultaneous glottal closure of the non-nasal [-cont] coda. Since oral stops do not entail simultaneous glottal closure while complete glottal closure does preclude nasality, we can say that (i) implies (ii) in a way that (ii) does not imply (i). It follows from the foregoing discussion that (i) is responsible for all of the other effects. If any phonetic feature is the realization of [*occl], it must be [* complete glottal closure at coda].

Glottal closure also explains the observation that -p -t -k have no audible release:

"pronounced without explosion" (Jones and Woo 1912:xi)

"have no audible release" (Wong 1940:Eng.6)

"In pronouncing the endings -p, -t, -k, (...) one must never let them 'go off' so strongly as to make an audible explosion. The sound is 'swallowed', as it is popularly described. In other words, a final -p, for example, is not pronounced as in a formally enunciated 'hope', but as in a carelessly answered 'Nope!'" (Chao 1947:22-3)

"All stop codas are unreleased." (Kao 1971:61)

"[T]he three voiceless stops, /-p -t -k/ are never released." (Dow 1972:161)

This is also in accord with O'Connor's (1980:140) description of a difficulty Cantonese speakers have when they speak English: "/p, t, k/ are not exploded in final positions."

Glottalization of final stops is spelt out in Egerod 1956:12 and Fok 1974:8, and is confirmed by an experiment done by Iwata (1985), who summarises that "[t]he basic laryngeal feature of final stops in Southern Chinese [i.e. Cantonese and South Min] is an adducted gesture of the glottis with an increased degree of supraglottic laryngeal constriction after the oral implosion."

8.2 Tones

The relativity of tonal distinctions is common knowledge. It is worthwhile, however, to distinguish between two different kinds of relativity pertaining to tonal distinctions.

The first kind of relativity has to do with the variability of the interval (which can be expanded or compressed) of vocal compass or pitch range, and the transposability of such compass or range. Thus, the exact F_0 value of such phonetic labels as "top/bottom of pitch range" varies from context to context; so therefore does the interval in physical terms. It follows from this kind of relativity that normalization of the interval and general pitch-level of any token of vocal compass is a prerequisite for any reference to tones, no matter whether the tones are handled in phonetic or phonological terms.

The second kind of tonal relativity concerns the interplay of contrasts imposed on and implemented within the normalized range of pitch. The interplay is relative in the sense that what matters is the maintenance of distinctions, not the exact positions in the normalized pitch-time space. It follows that it is the specific nature of the system of tonal contrasts at work in a particular language that determines, among other things, how many pitch-levels need be distinguished, and how pitch-orientation (falling, rising, even) is utilized distinctively.

The first relativity distinguishes between the physical/musical approach to pitch (F_0 /absolute pitch) and the linguistic approach to pitch (normalized pitch), and the second relativity distinguishes between the phonetic approach to tone (referring to the realization of tones in terms of exact positions in the pitch-time space) and the phonological approach to tone (referring to the interplay of contrasts for the maintenance of tonal distinctions). It is the phonetic characterization of Cantonese tones that we are here addressing.

The phonetic realization of tones is certainly not unrelated to the phonological status of the tones. For instance, since four pitch-heights are utilized for tonal contrasts, the pitch continuum is actually divided into four regions. The "cardinal relations" discussed in Section 3.3 give us some idea of the salient properties and the range of variability of the realization of each tone. But we would certainly like to have descriptions that are more concrete.

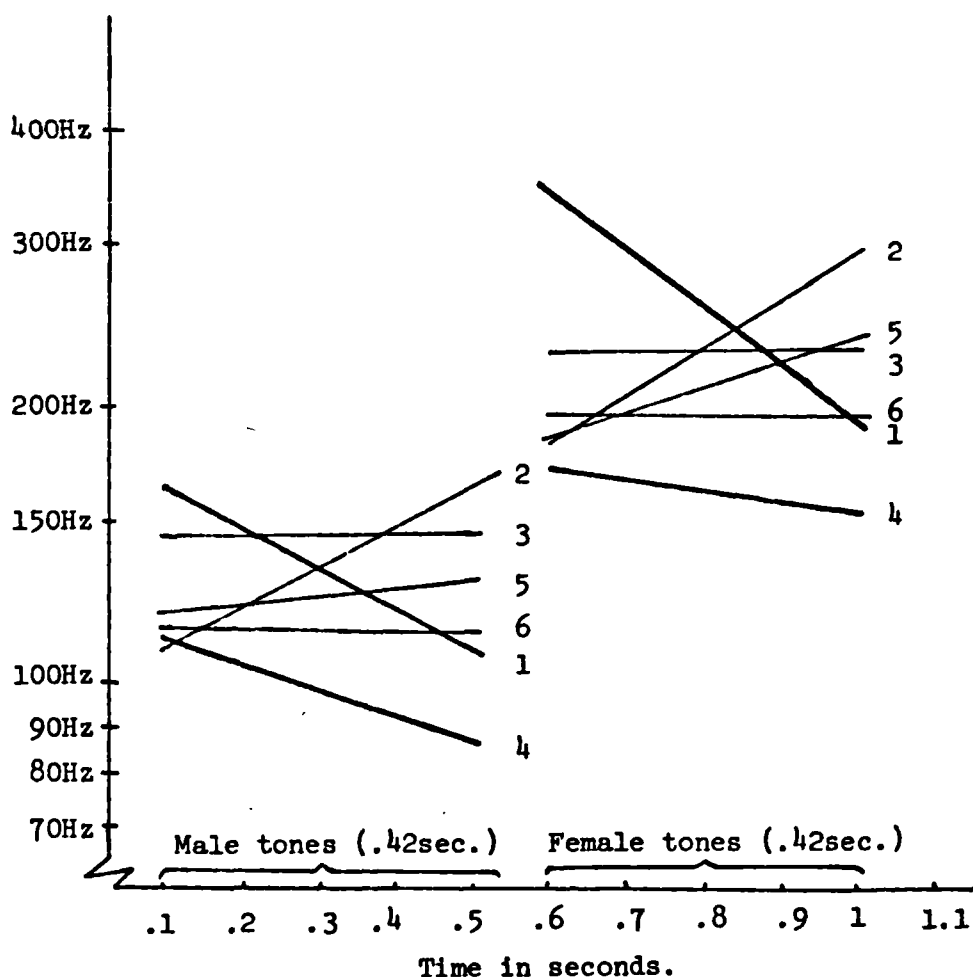
By dividing the (normalized) pitch range into four grades (instead of Y R Chao's five), we have characterized the six tones (including

their allotones) in Section 3.3 as follows:

[3] T1=44(HE)/42(HF), T2=24, T3=33, T5=23, T6=22, T4=21(LF)/11(LLE)

Except for HE, which is not represented in her experiments (though its existence is recognized), Fok's (1974) perception test done with the simplified synthetic tones¹ shown in [4] gives support to the characterizations in [3] above, as can be seen from the high success rate of tone identification shown in table [5].

[4]



¹ The synthetic tones were based on tones from a male speaker and a female speaker and were obtained by suppressing duration and intensity differences and by representing every tone in a straight pitch-time line.

[5] Confusion matrices with synthetic tones (column total 1080):

Male tones:	\Stimuli	T1	T2	T3	T4	T5	T6
	Subject\ Responses\						
	T1	871	0	526	18	7	6
	T2	3	1029	3	2	14	0
	T3	33	8	513	5	121	143
	T4	84	3	6	1014	0	13
	T5	14	37	14	4	890	62
	T6	75	3	18	37	48	856

Female tones:	\Stimuli	T1	T2	T3	T4	T5	T6
	Subject\ Responses\						
	T1	1062	29	1	4	0	6
	T2	2	886	6	7	41	10
	T3	5	16	925	14	20	54
	T4	2	6	10	979	5	33
	T5	4	141	22	8	1009	14
	T6	5	2	116	68	5	963

Since what matters is the maintenance of distinctions, both the characterizations in [3] and the representations in [4] must be over-specified in general. For the purpose of determining the range of realization for each tone, Vance's (1977) experiment on the perception of pitch-time graph in terms of Cantonese tones is valuable. Table [6], adapted from Vance 1977, shows the tone that receives the highest score of identification for every combination of beginning and ending F_0 in Hz in his experiment.¹

¹ Twenty responses were given for every stimulus, i.e. every combination of beginning and ending F_0 . The 'highest score' represents at least seven instances of identification, or 35% of the total. Most of the time the score is 10 or above, i.e. at least 50% of the total.

[6]	\Ending F ₀ →		92	106	119	133	154	178	206	245
	Beginning F ₀ ↓	\ Tone ↘								
	245		1	1	1	1	1	1	1	1
	206		1	1	1	1	6	1	1	1
	178		1	1	1/3	6	3	1	1	1
	154		4	6	6	3	3	5	5	5
	133		6	6	6	3	5	5	2	2
	119		6	6	6	5	5	5	2	2
	106		6	6	5	5	5	2	2	2
	92		6	5	5	5	2	2	2	2

By dividing the pitch continuum into four grades, from 1 (the lowest) to 4 (the highest), we obtain 16 different combination of beginning and ending pitch:

[7]	\Ending pitch		1		2		3		4	
	Beginning pitch ↓	\ Tone ↘								
4			1	1	1	1	1	1	1	1
			1	1	1	1	6	1	1	1
3			1	1	1/3	6	3	1	1	1
			4	6	6	3	3	5	5	5
2			6	6	6	3	5	5	2	2
			6	6	6	5	5	5	2	2
1			6	6	5	5	5	2	2	2
			6	5	5	5	2	2	2	2

The favoured tone (defined as the one scoring two or more times out of four) for each of the 16 combinations is as follows:

[8]	\Ending pitch		1	2	3	4
	Beginning\					
	pitch ↓	\ Tone ↘				
4			T1	T1	T1	T1
3			T1	T6/3	T3	T1/5
2			T6	T6	T5	T2
1			T6	T5	T2	T2

The correspondence between pitch-time graph and tone can be summarized as follows:

[9] a.	4X, 31:	T1
	33:	T3
	24, 14, 13:	T2
	12, 13:	T5
	34:	T1/T5
	32:	T3/T6

b. 32, 22, 21, 11: T6

[9a] corroborates both [4] and [5], and gives additional information about the range of realization for the tones concerned. In the case of T1, the range is actually predictable from the bundle of feature values that define the tone. Thus, [+extreme, +high, -rising] is compatible with all of 44, 43, 42 and 41; and the lack of another tone which is [+high, +falling] explains why 31 also tends to be perceived as T1. In the case of rising tones, Vance's experiment shows that the most important difference between T2 and T5 is one of gradient. The ending pitch, though highly correlated with gradient, is not the most reliable clue; thus 13 tends to be perceived as T2 and 34 as T5 (along with T1). The beginning pitch is even less relevant. In the case of T3, the fact that it has the largest number of contiguous tones (T1, T5 and T6 being respectively its [+extreme], [+rising] and [-high] counterparts) explains the strict requirement with respect to the shape of its pitch-time graph. On the other hand, because of the lack of any tone that is [-extreme, +falling], a slight fall is tolerable for T3 (and T6 as well for

that matter).

Vance's experimental results, however, show a strong favour for T6 and disfavour for T4, as can be seen from [7b]. Thus, while 32 and 22 are natural manifestations of T6, 21 and 11 are expected to be T4. There are a number of possible explanations for the phenomenon:

- 1) Despite the general disfavour for T4, it is nevertheless the favoured tone for the F_0 combination 154-94. Though the combination is classed as 31 in the simplified representation, 154 Hz is in fact closer to the lower than the higher end of the frequency range. Thus we can say that a very narrowly interpreted 21 is already favourably associated with T4.
- 2) LLE, or 11, may not be the favoured or permitted allotone of T4 for the particular morpheme (namely jew⁴ "oil") in citation form used in this experiment.
- 3) The subjects might be unfamiliar with synthetic tones.
- 4) T6 represents the default value of tone, and attracts identification when uncertainty arises.

Vance himself also points out the possibility that "something besides F_0 is involved in differentiating [T4] from the other tones". Despite the problem in Vance's study, the high success rate in the identification of T4 shown in [5] should serve to establish that 21, or less obviously 11, should be sufficient clue for the identification of T4.

Thus, except for the perception of 11, Vance's experiment is a fairly satisfactory indication of the range of realization for each tone.

Besides the pitch-time graph, which Fok 1974 and Ching 1981 have shown to be the necessary and sufficient condition for the identification of Cantonese tones, other concomitant properties have been reported or suggested. Thus, Vance (1977:105), in the course of explaining the discrepancy between the identification of natural and artificial T4, mentions that "some speakers (...) always produce [T4] with creaky voice". Ching (1981:70) also describes T4 as "often ending with creak". In the context of looking for something besides F_0 that differentiates [T4] from the other tones, Vance also suggests that "an intensity cue seems most likely", but he does not identify any such cue positively. I observe that the extreme pitch-level always couples with higher intensity, resulting in intensity contours in [+extreme] contoured

tone-shapes, i.e. HF, LF and T2. Thus, HF has a louder beginning, while LF and T2 have a louder end.¹ Duration has no correlation with tones whatsoever. Fok (1974:34-5) writes, "The duration of each of the normal tones (...) seems to be arbitrary. The length of the tones fluctuates but (...) in no regular pattern." Vance (1976:383), too, concludes, "Vowel length is apparently not involved in differentiating the Cantonese tones."

Despite the concomitant voice-quality and intensity cues, the importance of the pitch-time graph in the differentiation of Cantonese tones is overwhelming. As Fok (1974:87) notes, "when F_x information was removed (...) tones could not be named at all although duration and amplitude information remained."

8.3 Vowels

The combination of the five features for vowel, subject to implication relations and context-determined suspension of contrasts, gives rise to the set of distinctive vowels i:, y:/u:, ɛ:, œ:, ɔ:, a:, e, e/o and ɐ. We need the following rules to specify the occurrence of the complementary pairs e vs o ([10]) and y: vs u: ([11]):

$$[10] \begin{bmatrix} V \\ - \text{2nd M} \\ + \text{round} \end{bmatrix} \rightarrow [-\alpha \text{ back}] / \text{---} [\alpha \text{ cor}]$$

$$[11] \begin{bmatrix} V \\ +\text{high} \\ +\text{round} \end{bmatrix} \rightarrow \begin{cases} [+back] / [+ant] \\ [-cor] \text{ ---} \\ [-back] / \text{elsewhere} \end{cases}$$

[10] says that e occurs before -n and -j, and o occurs elsewhere. [11] says that u: occurs after labial onsets and y: occurs elsewhere. Having arrived at the eleven distinctly identifiable (but not necessarily contrastive) vowels, it is appropriate that we consider the quality of these vowels.

¹ Kratochvil (1968) has established convincingly the presence of intensity cues in the perception of Mandarin (his 'Modern Standard Chinese') tones. If my observation is correct, HF and T2 are then respectively similar to the Mandarin high-falling and high-rising tones with respect to not only pitch-time graph but also intensity contour.

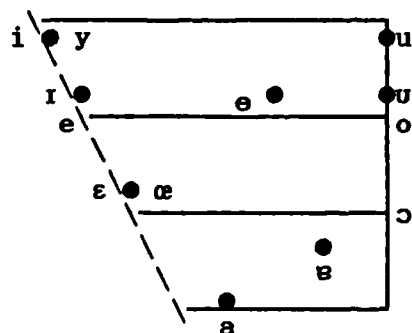
At the time of Jones and Woo 1912, the Cardinal Vowel system was not yet fully worked out (Abercrombie 1967:152). We therefore missed the chance of having Cantonese vowels represented in this system by the creator of the system. Fortunately the IPA 1949 includes a passage in Cantonese transcribed in the IPA, and the transcription, together with introductory notes, was done by D. Jones (J.C. Wells, personal communication). All vowels except *i*, *e*, *o* are given their "precise values", thus *a*=*a*·, *ɐ* is "Λ-like", *ɛ*=*ɛ*·, *ɔ*=*ɔ*·(=ɒ), *u*=*u*·, *y*=*y*·. In Jones' scheme *e* occurs in *ej* and *o* in *ow* only, and the two vowels presumably coincide with their Cardinal value. Jones transcribes *ɛŋ*/*k* as *iŋ*/*k* and *oŋ*/*k* as *uŋ*/*k*, but Jones and Woo (1912) describes *i* as "almost *e*" "before *k* and *ŋ*" and *u* as tending "towards *o*" "before *k* and *ŋ*".¹ As for *i* in general, it seems that it should have been *i*· as well in the IPA 1949, judging from the description of *i* as "resembling the English sound of *i* in *it*" in Jones and Woo 1912. *ɐ* is taken as *æ* in both works, but Lee 1985 shows that *æ* and *e* have non-overlapping F1 and F2 values in both Guǎngzhōu and Hong Kong Cantonese.

I have come across three representations of Cantonese vowels on the Cardinal Vowel quadrilateral, given in Yuán et al 1960:185, Huang 1965:28 and Fok 1974:7. Yuán et al's model is presumably Guǎngzhōu Cantonese, while Fok's is Hong Kong Cantonese. Huang's model is of unknown status. The same three vowels, namely *e*, *ɐ*, *o*, are left out in Huang's and Fok's representations. Yuán et al's "i" and "u" of course represent the *e* and *o* before *ŋ*/*k*. The three representations are reproduced below, followed by my own representation:

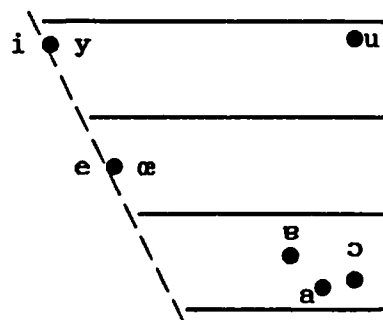
¹ Bear in mind, however, that the 'e' and 'o' may not be Cardinal.

[12]

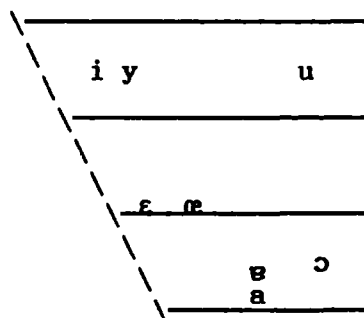
Yuán et al:



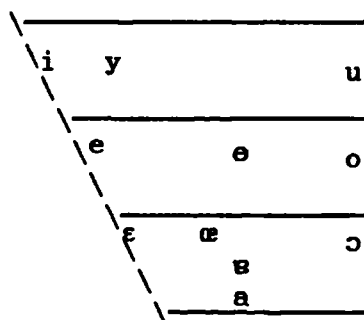
Huang



Fok:



This thesis:



My representation accords with the F1 and F2 values of the vowels given in Ching 1981:55, where e, e, o are again left out, and those given in Lee 1985, where though glide-checked vowels are excluded in general, these three vowels are nevertheless included by virtue of the rimes eŋ/k, en/t, and oŋ/k.¹

¹ The exclusion of glide-checked vowels is probably due to the influence of Jones and Woo (1912), who do not recognize these phonetic diphthongs as phonologically V+Cd. Another possible reason for the omission is the consideration that the first element of these glide-checked rimes is always an allophone of another basic vowel. This second reason also explains the omission of the vowels in eŋ, en and oŋ by some writers. This strikes me as hypostatizing 'phonemes': when realization rather than phoneme system is at issue, complementary

Tongue height is fairly variable, especially for y, u, a, ɐ. The variability of tongue height for a and ɐ results in a large area of overlap of their F1-F2 space for some speakers (Lee 1985). Tongue (front-/back-ness, on the other hand, is relatively stable. A "backness rule" applicable to non-high long vowels, implicit in Kao's (1971:29-30) description, whereby ɛ:, ɔ:, a: have a more advanced version before -n and -j and a more retracted version before -ŋ, seems to me arrived at through imagination and/or deduction rather than observation. If such a process exists at all, it has not "phonologized".¹ Besides, there is no reason why ɐ should be exempted from this alleged process.

Lip rounding is contrastive only for front non-low vowels. The non-contrastiveness of rounding has quite different effects on front low vowels and on back vowels. a and ɐ are realised unrounded as a rule. On the other hand, despite the implicational rule [+back] → [+round], the back vowels are often only "nominally" rounded. In general, the lower the back vowel, the weaker is the rounding tendency. However, it can be argued that a particular kind of "rounding" is present — labiodentality, especially for u after labiodentals, e.g. [fʷ]2 "bitter" and gw[ʷ]2 "guess". Compare the transcription [fv] by Dow (1972:164) and his comment that "/u/ is not so rounded as that in Peking dialect"(p.160).

Apart from the above-mentioned variation, no other notable variability exists in the realization of Cantonese vowels.²

8.4 Codas

distribution is irrelevant and any distinct realization should be taken care of. My comments apply to formant measurements and representations on the Cardinal Vowel quadrilateral alike. Fok (1974:8), for instance, observes that '/ɛ/ and /ɔ/ when followed by a high vowel become slightly higher themselves, resulting in /ei/ and /ou/', but she does not include e or o in her representation of vowels on the C.V. quadrilateral.

¹ Refer to Section 7.2.1.1 for the idea of phonologization.

² Jones and Woo (1912:xii), however, mention that in the syllables dzi, tsi, si, 'the vowel is accompanied by considerable friction, and consequently resembles a very fully voiced z'. This allophone of i: has disappeared from Hong Kong Cantonese completely, and seems to have disappeared from Guǎngzhōu Cantonese too, as it is mentioned in Yuán 1960:185 only anecdotally, viewed as something alleged only.

The typical place of articulation of the various codas is as follows:

m/p:	bilabial
n/t:	apico-postdental
ŋ/k:	dorso-prevelar ¹
j:	dorso-palatal
w:	simultaneously endolabio-dental(/bilabial) and dorso-velar

As far as Hong Kong Cantonese is concerned, the labiality of -w is not likely to be in the form of bilabiality (which entails the rounding or protrusion of the lips), unless it follows a rounded vowel.²

The occlusive codas -p -t -k have simultaneous complete glottal closure. these codas are thus phonetically [ʔp ʔt ʔk]. In the case of -k, the velar articulation is sometimes lifted, resulting in a simple glottal stop [ʔ].³ Realization of -k as [ʔ] has been reported in Bauer 1979:133.

The oral constriction for all the [-cont] codas is complete closure. Of this class of codas, the non-occlusive ones of course differ from the occlusive ones in that lung air is escaping first through the vibrating vocal folds and then through the nasal cavity.

The dorsal constriction for [+cont] codas, i.e. the constriction for -j and for the velar articulation of -w, is vocoidal, which does not necessarily result in an approximant. The full-moraic -j and -w are realized as the closest kind of vocoids, namely [i] and [ɰ/u]. As such they are of course approximants. The semi-moraic -j and -w, on the other hand, in general have opener constriction (in the dorsal region)

¹ 'Prevelar' is a zone of upper articulator recognized in Canepari 1983:79. It is between palatal and velar. If such a zone is not recognized, such as in the scheme of localization adopted in Catford 1977 or Wells 1969, the place of articulation should be velar rather than palatal, thus [ŋ] and [k̟].

² See Section 8.7 below for further details.

³ When Jones and Woo (1912:xi) remark, 'Final t and k are often replaced by the "glottal stop" (...), especially when the next word begins with h, j or w,' they seem to be confusing glottalization with glottal stop: [ʔt, ʔk, ʔ] are all identified as glottal stops with the lingual closure undetected, while [ʔp], thanks to the conspicuously observable bilabial closure, is identified as plain bilabial. [ʔ] in fact alternates only with [ʔk], not with [ʔt].

than the full-moraic counterparts. The "trochaic" diphthongs, i.e. rimes checked by semi-moraic -j or -w, in general involve only a small amount of tongue movement, just enough to suggest a narrowing towards [i] or [ɰ/u], but seldom reaching the target. As can be expected, the opener the preceding V, the opener -j and -w tend to be. Thus a:j can be [a:ɛ̃] and a:w can be [a:ʏ̯].

The labial articulation of -w is in the form of close approximation or higher degree of stricture. When the labial articulation is labiodental, friction is permissible. Friction is unlikely, however, when -w is semimoraic and preceded by a non-high vowel, i.e. in a:w and ɛ:w. Compare ɛw, ow and i:w, for which the labial articulation may be in the form of a weakly frictional [v].

The zero coda is realized as the prolongation of the preceding V, as the moraic organization of the syllable provides for and indeed predicts.

8.5 Onsets

The place of articulation of the various onsets is as follows:

gw kw	simultaneously bilabial/labiodental and dorso-velar
w	bilabial/labiodental and (optionally) simultaneously dorso-velar
f	labiodental
b p m	bilabial
d t l	apico-postdental
dz ts s	lamino/dorso-prepalatal ¹
j	dorso-palatal
g k	dorso-(pre)velar
ŋ	dorso-(pre)velar/pharyngeal/glottal
h	homorganic with the following sonorant sound

¹ See the next section (8.6) for further details.

The manner-of-articulation correlates of the various feature-defined classes of onsets are as follows:

[-cont] → oral stops

[-cont, -voic] → long voicing lag, i.e. voiceless aspirated

[-cont, +voic] → short/no voicing lag, i.e. unaspirated

[+cont, -voic] → voiceless fricatives¹

[+cont, +voic] → voiced approximants

The simple pattern given above is subject to the following refinement and variation.

gw- and kw- differ from other [-cont] onsets in that they involve two places of articulation simultaneously^{ly}. Though the labial articulation is of lower-ranking degree of stricture than the velar articulation, it is nevertheless not necessarily less important than or secondary to the latter, since it lingers for a while even after the velar stop has been released. The labial articulation is an approximant or is weakly frictional.

The labial articulation in w- is even more important than that in gw- and kw-, because for w- the labial constriction is of equal or higher rank than the velar constriction. As in gw- and kw-, the labial articulation is an approximant or is weakly frictional. Labiodentality and friction of w- have been reported by Gão (1980:2).²

The [-cont, +voic] onsets have been customarily characterized as voiceless unaspirated. We have avoided the awkward association between [+voic] and phonetic "voicelessness" by reference to VOT (voice-onset time). There are independent justifications for this course of action.

First, Lisker and Abramson (1964:387) have argued eloquently for the relevance of VOT to both voicing and aspiration:

[W]e may define the amount or degree of voicing of a stop as the duration of the time interval by which the onset of periodic pulsing either precedes or follows release. In thus giving up the absolute

¹ The characterization of [h] as a fricative is conventional. Cf. the last footnote.

² Yuán et al (1960:183) describe w- as a bilabial and velar voiced fricative. The description is unsatisfactory in that it suggests that friction exists at the velar as well as the labial articulation, which is not the case.

definition of the term "voiced stop" with which we begin, we are free to say that a difference of voicing not only separates voiced from voiceless stops, but that it equally well distinguishes aspirated from unaspirated stops, where the latter are both commonly called voiceless. The noise feature of aspiration, instead of being considered coordinate with voicing, is then regarded simply as the automatic concomitant of a large delay in voice onset.

Second, despite the accepted characterization of [-cont, +voic] as "voiceless", there are times when these onsets have voicing lead, i.e. positive VOT value. Thus, Vance (1976:383) reports that in certain contexts the affricate in dzi:⁶ ("word/character") "was often voiced through" in his sample sentences. Impressionistically, I also observe that [-cont, +voic] onsets when non-initial often have voicing lead in casual speech. The following are some of the notable examples:

- [13] a. hɛm⁶ba:⁶la:ŋ "all"
 b. a:³ba:⁴ "papa"
 c. gɔ:²di:¹ "those"

The two considerations above together suggest that both voicing lead and short voicing lag are manifestations of [+voic] when [-cont]. As a first approximation, I formulate [14] to account for the different realizations:

- [14] [+voic] → (short)¹ voicing lag / #_____
- [-cont]

Rule [14] says that some kind of "juncture" # is responsible for the "devoicing" of [-voic, +cont] onsets. I am not entirely clear what constitutes # exactly, but pause and word boundaries imply its presence.² Note that dzi:⁶ mentioned by Vance and di:¹ in [13c] are used as a suffix and a:³ in [13b] and gɔ:² in [13c] used as a prefix, while [13a] contains no morpheme boundary at all. A pause can be inserted before the stops in question in these examples, and the post-pausal [+voic] stops will then have no voicing lead.

Both Lisker & Abramson (1964) and Clumeck et al (1981) show that

¹ The specification 'short' will be rendered redundant presently.

² See Chapter 10 for further discussion on #.

the difference between [+voic] and [-voic] in Cantonese [-cont] onsets is one of short vs long voicing lag. The mean VOT in ms for the six simple Cantonese initial stops in the two experiments is given below:

[15]		b	d	g	p	t	k
	Lisker & Abramson	9	14	34	77	75	87
	Clumeck et al	10	10	26	74	83	91

The lack of tokens with voicing lead is attributable to the fact that both experiments are concerned with initial stops. In conjunction with what we know about onsets with voicing lead, these experiments show that a VOT of about 40 ms is the boundary between [+] and [-voic] for Cantonese stops: a later onset means [-voic] while an earlier onset, lead or lag, constitutes [+voic].

The actual realization of [+cont, +voic] is quite varied. m- and ŋ- are nasals, but ŋ- can also be realized as [ʔ] or the voiced pharyngeal fricative [ʕ]. l- is typically a lateral approximant, often nasalized, i.e. [l̃],¹ but it is well known that sometimes and/or for some lexical items some people use [n], as a result of a sociolinguistic resistance to the merger of l- and n-. j- varies between approximant and fricative. Friction in j- has been reported by Yuán et al (1960:183) and Gāo (1980:2). The manner of articulation of w- has been described above. Like l-, both j- and w- are often nasalized.

8.6 The place of articulation of sibilants

Of all the onsets, the place of articulation of the sibilant series dz, ts, s is the most variable and its description most varied and controversial. The sibilants have been variously notated by different writers as follows:

[16]	Jones and Woo (1912:xi-xii)	ts ts' s
	Chao (1947:28)	tɕ tɕ' ɕ
	Most linguists in China, e.g. Wang (1936-7:641)	tʃ tʃ' ʃ

¹ The nasalization is independent of the pre-merger distinction between n- and l-, which is still prescriptively and mythically maintained nowadays. Nasalization of l- has been mentioned by O'Connor (1980:140) in passing.

Despite the different notations employed by Jones & Woo and Chao, the verbal descriptions of the sounds by them are not really incompatible. For Jones and Woo, *ts* resembles a sound intermediate between the sounds of *ch* in *chin* and *j* in *jam*"(p.xii), and *ts'* "bears a certain resemblance to the English sound *ch* in *chin*"(p.xi). They ask,

Should [*ts'*, *ts*] be written *ts'*, *ts* or *tʃ'*, *tʃ*? The actual sounds appear to vary between these alternative values. (p.viii)

Chao, on the other hand, gives specifications such as "between *chat* and *adʒə*", "between *it's hot* and *such heat*" and "between *she* and *sell*" on the *List of Values of Initials* (p.28). In addition he writes,

[A]n English-speaking student should guard against the tendency to round or protrude the lips (except of course, with words having the rounded vowel *ü*). (p.19)

Jones and Woo have good reason for not using "*ʃ*": at the time when they wrote, the symbol was not yet in use (J.C. Wells, personal communication)! And there is good reason for them to use "*s*" in particular: according to them, the Cantonese *s*, unlike the affricates, is simply "as in English".(p.xii)

Phonetic notations are seldom decided on free of pragmatic considerations. Chao's use of "*ʃ*" (as opposed to "*f*") serves well to remind English-speaking learners not to round or protrude the lips. On the other hand, descriptions of Cantonese sounds in Chinese are anxious to keep them apart from Mandarin sounds. Since "*s*" has been used for the Mandarin dental sibilants and "*ʃ*" is illustrated by the Mandarin prepalatal in the IPA 1949:12, there are good pragmatic reasons for Chinese linguists to use *f* for the Cantonese sibilants. However, one easily hypostatizes the relationship between a symbol and a sound that it is recurrently used to stand for. We find just this sort of hypostatization with respect to "*f*" and the Cantonese sibilants. Thus Cen (1982:23-4) writes:¹

[*f*] [*ʃ*]. "Blade fricatives". To produce these sounds, put the tongue-tip against the lower gum, raise the tongue-blade near to

¹ My translation and my emphasis.

the alveolar-prepalatal region, leaving a narrow channel; slightly round and protrude the lips; (...). [ʃ] is voiceless, as [ʃ] in Cantonese [fan] "mountain" (...)

Cen's description is inadequate on three counts. First, there is no intrinsic relation between the IPA symbols ʃ, ʒ and lip rounding. Second, despite and in contrast to a certain degree of lip-rounding in English ʃ and ʒ, Cantonese sibilants are not intrinsically rounded, as Chao's warning above clearly shows. Third, Cen's description as regards the position of the tongue-tip, though in accord with Cantonese sibilants, is not a common property of [ʃ, ʒ] in general. According to Catford (1977), the tongue-tip for [ʃ, ʒ] either directly forms a stricture with the "postalveolar" region (p.152), or rests some distance retracted from the lower teeth, leaving "a little cavity" in between, when the blade of the tongue serves as the active articulator.(p.154-5)

Adequate description of sounds consists in precise verbal description rather than the mere choice of symbols from a limited set. Jones & Woo and Chao are already on this road when they compare the Cantonese sounds with the English. Also in connection with verbal description of Cantonese sounds S Cheung (1972:2) quotes the words of Y.R. Chao which he has obtained through personal communication (my translation).

Compared with the English s, dz, ts, these three sounds are in fact somewhat ʁ-coloured with respect to place of articulation, for I feel that the Pekinese s ʃ ʁ are a very front s, a very retroflex ʃ and a very flat-tongued ʁ. The English s, etc, is similar to the Pekinese s, etc, only a little retracted. (...) As for the corresponding sound in Cantonese, according to my observation, what is most common in Guǎngzhōu City is a place of articulation between those of s, ʃ and ʁ:

*ʁ
s *
*ʃ

That is why speakers of other Chinese dialects keep hearing it as "hushing". Only the Xiguān accent of Cantonese uses a more apical sound.

Chao's characterization serves as a perceptual, contrastive description. Yet it is still not a positive description in articulatory terms. For the latter purpose Dow (1972:157) has much to offer:

These three Cantonese sounds are generally articulated by placing the tip of the tongue against the back of the lower front teeth. The area between the blade and the front of the tongue is raised against the region between the alveolar ridge and the hard palate. They have a sound value between apico-alveolars and fronto-prepalatals.

Dow sets a good example of characterizing the place of articulation in terms of both active and passive articulators. He spells out, among other things, that the tongue-tip is not so retracted as to leave a post-dental cavity as is typical of "palatoalveolars", e.g. the English /ʃ/. Dow's last sentence suggests that the Cantonese sibilants lie (perceptually) between alveolars and "alveolopalatals", but are neither of them. In a sense, then, the characterization of Cantonese sibilants in terms of s, ʃ, or ɹ is a matter of definition and/or classification. However, convention gives preference to one choice or another. There are two other parameters we have not considered, namely (i) the cross-sectional shape of the constriction, and (ii) the length of the constriction along the midsagittal axis of the oral tract. As Brosnahan and Malmberg (1970:102-3) note:

The characteristic acoustic qualities associated with each of these articulatory variations are sufficiently prominent to dominate differences of resonance introduced by small changes in the position of the constriction.

The constriction in [s] is typically more "grooved" and shorter than that in the Cantonese sibilants in question. We therefore have good reason not to regard the Cantonese sibilants as alveolars. The choice between ʃ and ɹ, on the other hand, is more a matter of personal inclination. Catford (1977:155), for instance, recognizes

an anomalous [ʃ] with a regular lamino-postalveolar channel, but with the point and rim of the tongue in contact with the rim of the lower teeth. (...) described as 'between [s] and [ʃ]' or 'hissing-hushing' (...).

If we, following Catford, recognize this "anomalous [ʃ]", then the Cantonese sibilants fall neatly into this category. Otherwise ɹ is a better candidate. The matter, however, is complicated by the various conceptions of the symbol "ɹ" (together with its voiced counterpart) or the term "alveolopalatal":

Primary articulation between "front of tongue and hard palate", secondary articulation between "blade of tongue and alveolar ridge". (Wells 1969:394)

"dorso-alveolar" (Brosnahan and Malmberg 1970:102)

"palatals with an alveolar modification, that is, dorso-prepalatal fricatives"¹ (Catford 1977:158)

"(Lamino)prepalato-(pro)labial"² (Canepari 1983:37)

Compared with the Mandarin alveolopalatal, which along with the Polish counterparts serve to illustrate and indeed define the term and the symbol "ɲ" in the IPA 1949:12, Cantonese sibilants are a little fronter. This small difference is easily misleadingly exaggerated by the fact that the Mandarin alveolopalatals are always followed by a high front vowel. If we compare the Cantonese si: with the Mandarin ɕi, the difference, although still perceptible to trained ears, is very much reduced. The characterization of the Cantonese sibilants as palatoalveolars or alveolopalatals is indeed a matter of personal inclination. While the symbol "ʃ^s" found on the IPA chart 1979 may serve well to represent the sounds in question, I favour "ɲ", etc, if only for typographical convenience.

A characterization of the typical realization of Cantonese sibilants, however, is not the end of the story, for variant forms exist. Jones and Woo (1912:xii), for instance, describe the Cantonese s as "as in English", suggesting that the s has a different place of articulation from the affricates.

Kao (1971:62) draws our attention to a context-determined variation:

Before the front vowel /i/, /c c' s/ are articulated in dental-alveolar position; before /æ/ and especially before /æŋ/, in a position between alveolar and pre-palatal; before /y/, in the same position as before /æ/, but with lip rounding.

¹ Note the possible different interpretations of these two characterizations Catford takes to be equal: 'prepalatals' are presumably fronter than 'palatals'.

² Canepari's 'prepalatal' does not refer to the region between (post)alveolar and palatal, as the term is usually taken to mean, but is a cover term incorporating both 'alveolar' and 'postalveolar'. As for '(pro)labial', it presumably refers to lip-protrusion. The version without lip-protrusion he symbolizes as 'ɕ' (with "ɲ" on the left) and 'ɕ'.

However, to the extent that y and æ are no more palatal than i, the alternation she describes is difficult to explain. In contrast, the variation described by Hashimoto (1972:89) is phonetically more plausible:

The sibilants are usually pronounced with dental articulation in the author's idiolect: other speakers vary from a dental (hissing sound) to a palatal articulation (hushing sound). Often these sibilants are pronounced with some degree of palatalization when they precede high front vowels.

And her own discrimination between s and dz/ts with respect to their palatalization — "The author sometimes palatalizes [ts] and [ts'] under such conditions, but scarcely [s]." (p.120) — parallels Jones and Woo's discrimination between s and dz/ts with respect to place of articulation.

If Hashimoto's description is correct, then at least for some speakers, alveolars are used when the sibilants are not followed by high front vowels.¹

I observe another kind of variation in the pronunciation of sibilants in Hong Kong. For some of those who are able to produce the English palatoalveolars correctly and are anxious to maintain this correct pronunciation, there is a tendency to substitute the English palatoalveolars for the usual realization of Cantonese sibilants when these are followed by rounded vowels, especially within a wider context of English-Cantonese code switching and code mixing.² In general, the rounder the vowel, the more likely is the substitution. Thus, the process affects the sibilants followed by y:, æ:, e, and, to a lesser extent, o and ɔ:.³ As we shall see in the next section, the roundedness of onsets assimilates to that of the V. What is taking place is that the speakers equate the rounded Cantonese sibilants (basically alveolopalatals) with the English palatoalveolars, pronouncing them with

¹ 'Dental' is the actual word used by Hashimoto. She apparently uses 'dental' loosely to include alveolar: in another context (p.88) she characterizes the sibilant series as 'alveolars/palatoalveolars'.

² This variation was first reported by Yeung 1980. However, she does not provide any conditioning factors for the variation.

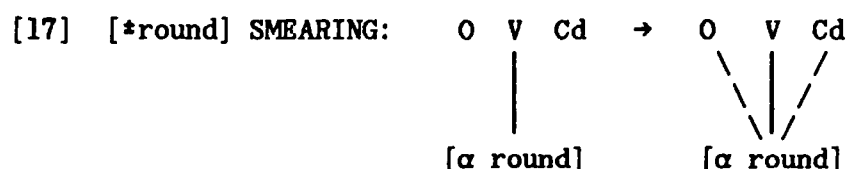
³ The sibilants do not co-occur with [u:]. See Section 8.3, especially rule [11].

the tongue-tip resting some distance retracted from the lower teeth, leaving "a little cavity" in between as described by Catford (1977:154-5) of the typical palatoalveolars. Seen in this light, the alternation described by Kao cited earlier becomes more reasonable: the alternation is one between rounded and unrounded sibilants. Rounded sibilants, no matter whether meant to be alveolar or alveolopalatal in normal circumstances, are palatoalveolarized (in the English way) rather than palatalized or alveolopalatalized.

To recapitulate, the Cantonese sibilants are typically alveolopalatal. Some people use the lamino-alveolar counterpart, and tend to (alveolo)palatalize them before high front vowels. The affricates have a higher tendencyⁿ to be realized as alveolopalatals than the fricative. For some people in certain contexts, English-like palatoalveolars, always with lip rounding, are used instead of the more usual realization.

8.7 Lip-rounding harmony

Lip-rounding harmony operates in Cantonese, and the domain of the harmony is the syllable. Note that [\pm round] is a feature of V, not of O or Cd. For Cd, -w and -m/p make up a [+lab] class. For O, the labials are characterized as [+ant, -cor]. The [+round] vowels include y:, æ:, e, u:, o, ɔ:, of which the front ones involve a higher degree of lip-protrusion^u/rounding. Whatever the degree of lip-protrusion/rounding, the feature [\pm round] spreads to the O and the Cd. The process can be depicted in the mode of presentation of autosegmental phonology as follows:



A consequence of this regularity is that each onset or coda has a [+round] and [-round] version, irrespective of whether it is intrinsically labial itself. In this perspective, the symbols "j" and to a lesser extent "w" are wrongly overspecified. In particular, when -j is full-moraic and follows a very rounded V, i.e. in the rime ej, the rounding of -j is most notable; hence the practice of representing this particular

rime as "ey" (or "øy").

For non-labial onsets and codas, this (non-contrastive) difference in lip-rounding is readily understandable. Compare, for instance, the d and n in dy:n and di:n. It is remarkable that the difference applies to labials as well, including gw-, kw-, w- and -w which are often misleadingly characterized as "rounded". Consider the following table:

[+round]		[-round]
MORE LIP-PROTRUSION	LESS LIP-PROTRUSION	
gw-	gwɔ:k ³ "country"	gwəj ¹ "tortoise"
g- gæ:3 "a saw"	gɔ:k ³ "horn"	gəj ¹ "chicken"
w-	wɔ:1 "barking"	wɔ:5 "rotten"
f- fæ:1 "fur"	fɔ:2 "fire"	fej ⁴ "fat"
m-	mɔ:4 "to grind"	mi:n ⁶ "face"
p- pæ:6 for "percent"	pow ³ "shop"	pi:n ¹ "edit"
-w	gow ¹ "tall"	gi:w ³ "to call"
-m	bom ⁴ "Boom!"	ji:m ⁴ "salt"
-p	bop ⁴ "heart beat kind of sound"	ji:p ⁶ "leaf"

We can see from the table that labials, like non-labials, have [+round] difference. The examples with g- are given for comparison with gw-items. Note, in particular, the [+round] gæ:3, with a high degree of lip-protrusion, and the [-round] gwəj¹.

CHAPTER 9: VARIATION

So far in this thesis Cantonese has been treated more or less as a unitary system.¹ This is of course the result of idealization, which is the starting point for any fruitful linguistic description. Variation, however, exists in the sound system of Cantonese, as it is bound to exist in any sub-system of any language. A phonological description is therefore incomplete unless variation is also taken care of.

Variation may be (1) regional, (2) chronological, (3) social or (4) stylistic. As far as pronunciation is concerned, regional variation, and to a lesser extent chronological and social variation, are often referred to as "accents". The study of accents from a synchronic, descriptive perspective is relatively new, though by now well established as an independent discipline. In a wider perspective, the study of accents can be regarded as a particular branch of an emerging area of study, namely the "'variationist' approaches to linguistic analysis", i.e. "studies which focus on the variable and dynamic aspects of language, in contrast to the categorical and essentially static orientation of modern theoretical linguistics"(Harris 1985:1). While the study of accents by definition deals with pronunciation alone, there is no reason why variation in relation to pronunciation cannot be singled out as an area of study, namely phono-variational study, which deals with all four kinds of variation listed above. The present chapter is devoted to precisely this kind of phono-variational description of Cantonese.

9.1 A taxonomy of pronunciation variations

There can be at least two ways of classifying variations. "Regional, chronological, social, stylistic" are attributes referring to non-linguistic conditions for variation. On the other hand, from the viewpoint of the effect of the variation on the language as a system, variations can be classified along a different dimension. It is this latter kind of classification that we are addressing in this section.

¹ Except in the account of realization in the last chapter where alternative pronunciations are listed.

If only for convenience of exposition, a language can be decomposed into two main components, namely the lexicon and the remainder, which is believed by many to be made up of a set of rules. When it comes to pronunciation variation, the dichotomy gives rise to two distinct kinds of variation. One kind affects the lexicon only, having no immediate effect on the phonology (e.g. the alternative pronunciations /'i:ðə/ and /'aɪðə/ for the word "either"), while the other kind always has some immediate effect on the phonology, whether or not the lexicon is also affected.

Phonological variations, in turn, can be dichotomized into (i) variations which affect the low-level realization of abstract, distinctive phonological entities, without effect on the play of contrasts in the sound system, e.g. the realization of /əv/ as [əv, ɐv, ov, ɔv] or the unrounded counterpart of [əv] within RP (Wells 1982), and (ii) variations that do affect the play of contrasts in the sound system.

Compare the following remarks by Trubetzkoy (1931):¹

Phonic differences between two dialects may be of three kinds: they may concern the phonological system, or the phonetic realization of the various phonemes, or the etymological distribution of the phonemes in words. Accordingly we shall speak of phonological, phonetic, and etymological differences between dialects.

Here the taxonomic similarity between accents and pronunciation variations is obvious. Trubetzkoy's tripartite division, though rigorous and inspiring, has not really exhausted all kinds of accent difference. Wells (1982:73, 75-6) makes the classification more complete by introducing a fourth kind of difference, namely phonotactic differences (e.g. between rhotic and non-rhotic accents of English).

In both Trubetzkoy's and Wells' scheme of classification, purely lexical differences can be singled out as having no bearing on the phonological component of the grammar of a language. It is for this reason that terms like "accent phonology" and by analogy "variation phonology" are not without problem if we want to include

¹ English translation taken from Wells 1982:73.

Etymological-distributional	Lexical-incidental	
Phonetic	Realizational	
(Not provided for)	Phonotactic	Combinational
Phonological	(Phoneme-)systemic	(Paradigm-)systemic

Of the four types of pronunciation variation, that of realization has been dealt with in the last chapter, when we deal with realizations in general. Realizational variations will therefore not be described in this chapter, though their correlation, together with the correlation of other variations, with non-linguistic variables will be dealt with in the final section (9.5) of this chapter.

9.2 Systemic variations

Systemic variation in Cantonese phonology involves the contrastiveness or otherwise of *n-* and *l-*, *ŋ-* and *ø-*, *-ŋ* and *-n*, T3 and T5, and the tone shapes HE and HF. These will be discussed one by one in this section.

9.2.1 [*±* merger of *n-* and *l-*]

A distinct onset *n-* is not recognized in our reference description. An onset *n-* in addition to *l-* surely existed. Chao (1947:18) asserts that "[a]bout one out of four persons in Canton has no initial *n*, and pronounces an *l* in words beginning with *n* for other speakers". This is probably the first record of a merger in progress. To represent the current situation where the majority of speakers in Guǎngzhōu and Hong Kong alike do not contrast *n-* and *l-*, I take *n-/l-* as one distinct onset rather than two. To acknowledge the fact that *n-/l-* is realized as *l-* (more exactly [l̥]) most of the time, I have notated the onset as "l-". The *n-* vs *l-* contrast, however, is prescriptively and/or pretentiously maintained. There are two ways of handling the minority cases (in terms of number of speakers and/or tokens) where the contrast is maintained. We can either say that all those who manage to maintain the contrast do so because of their knowledge of some other Chinese dialect (or

dialects), i.e. because they are bilinguals, or say that a systemic variation exists, differentiating two kinds of speakers qua Cantonese speakers, or two accents of Cantonese.

The first option, i.e. the bilingualism explanation, is the one I prefer. As far as the n- vs l- contrast is concerned, while non-merger is the dominant case merger is not uncommon among Chinese dialects; and merger and non-merger dialects exist side by side within the same dialect group. For example, Lánzhōu (of the Northern dialect group), Fúzhōu (of the Mǐn dialect group) and Sǔi'ān (of the Wú dialect group) are merger dialects while most other dialects in these dialect groups are non-merger dialects (Yǔyán Yánjiū-suǒ 1958). On the other hand, the lexical incidence of n- and l- is highly correlated across all non-merger dialects. In particular, the prescribed lexical incidence of n- and l- in Cantonese is virtually the same as that in other dialects which Cantonese speakers have the most contact with, including Mandarin (of the Northern group) and other dialects of the Yuè group. It is because of this pan-dialectal lexical incidence of n- and l- that we can attribute the maintenance of such contrast by the minority of speakers of Cantonese to bilingualism. This is of course ultimately an empirical issue. If it can be shown that there are monolingual speakers of Cantonese who maintain the contrast, then the bilingualism explanation should give way to the systemic variation explanation. Another possibility is that both bilingualism and systemic variation contribute to the present state.

Since at least for the merger speakers n- and l- are free variants, "hypercorrection" can be detected from time to time. Thus, not only is the prescribed n- pronounced l-, which is the usual though slightly stigmatized realization, but the prescribed l- is sometimes realized as n-.¹

9.2.2 [η - and \emptyset -]

Like the contrast between n- and l-, that between η - and \emptyset - is not

¹ The existence of this kind of 'hypercorrection' has consequences on the way we interpret the realizations of both the prescribed n- and the prescribed l-. See Section 9.5.2 below for details.

recognized in our reference description. But unlike n- and l-, the ŋ- vs Ø- contrast has never been clear-cut. For instance, while recognizing the contrast, Chao (1947:20-1) writes:

There are two types of usage in regard to words beginning with the open vowels. Except for interjections, particles, and the proper noun prefix [a:]³, which begin with an open vowel for all types of speakers, about one fourth of the speakers of Cantonese pronounce this group of words with an open-vowel beginning while the remaining three fourths pronounce them with initial [ŋ-].

This description, however, is not the whole story. In Chao's system, ŋ- and Ø- are not contrastive for either the ŋ- only or the ŋ- cum Ø- usage: if Ø- is ever used, it is in complementary distribution with ŋ-: Ø- with T1-3 and ŋ- with T4-6.¹ The apparent complementary distribution is the result of historical development which does not concern us here. For those who do maintain the ŋ- vs Ø- distinction the complementation is upset by such forms as ŋew¹ "hook", which is (diachronic-)derivationally anomalous and would be gew¹ in the normal course of derivation, and ɲa:m¹ "correct", which is a "new" item, i.e. having no etymon in MC.² There are, therefore, (theoretical) minimal pairs such as ŋew¹ "hook" and gew¹ (a surname). The realization of these two theoretical onsets exhibit a whole range of possibilities. The following table presents all these possibilities:

¹ Interjections, particles and the proper noun prefix [a:]³, which Chao cites as exceptional, are excluded from our consideration. Interjections are different from other morpho-syllables in that they never participate in word-formation processes and are carriers of intonation, not tone. Particles are peculiar in that they are enclitics and the noun prefix [a:]³ in that it is a proclitic. Particles, which are phrase-final, are also intonation carriers to a certain extent. The three classes of items make up a closed set. Moreover, these items have no phonetic onset at all, unlike Ø, which is usually realized as [ʔ].

² Other "new" items of this kind include ɲen¹ "small", ɲa:k¹ "deceive" and ɲep¹ "to utter".

[3]	<u>ŋ- + T4-6</u>	<u>ŋ- + T1-3</u>	<u>Ø- + T1-3</u>
a. Difference maintained:			
i. Contrastive	ŋ	ŋ	Ø
ii. Strictly complementary	ŋ	Ø	Ø
iii. Complementary (optional: Ø→ŋ)	ŋ	Ø/ŋ	Ø/ŋ
iv. Complementary (optional: ŋ→Ø)	ŋ/Ø	Ø	Ø
b. ŋ- only	ŋ	ŋ	ŋ
c. Ø- only	Ø	Ø	Ø
d. Free variation	ŋ/Ø	ŋ/Ø	ŋ/Ø

Cheng (1968:19) has similar descriptions of the range of possibilities:

- (a) some speakers keep the distinction;
- (b) some pronounce both sets with initial ŋ;
- (c) some drop initial ŋ from the system and pronounce both sets with the zero initial; and
- (d) some may even pronounce words of Yīng group [i.e. with Ø- theoretically] with initial ŋ and words of the Yí group [i.e. with ŋ- theoretically] with the zero initial, just the reverse of what is considered standard.

The correspondenceⁿ between her (a,b,c) and mine is obvious. However, her loosely worded (a) is ambiguous in that it can refer to any of the four situations in my (a). The "distinction" categorically constitutes "contrast" only in the case of (i). (ii) is a neat case of complementary distribution and ŋ- and Ø- are therefore collapsible. However, T1-3 and T4-6 as natural classes only have diachronic motivation; in the synchronic system they are not independently motivated. The alternation in (ii) is therefore unlikely to be a stable situation, i.e. it is more likely indicative of contrast between ŋ- and Ø-, with historical and accidental gaps in their combination with tones, rather than of merger. (iii) and (iv), on the other hand, seem to be particular cases of (b) and (c) respectively, with remnants of the effect of (ii). On the face value of her words, her (d) appears to me empirically false: a systematic reversal of (i), i.e. an automatic "swap" of the theoretical ŋ- and Ø-, is unheard of. Nevertheless, a consequence of my (d) is that ŋ- sometimes occurs when Ø- is expected and vice versa. Her (d) may be interpreted to refer to such a situation so that it is compatible with my (d).

A η - vs \emptyset - contrast is not recognized in our RD. This is because of the small proportion of speakers who adopt position (i) or (ii). To provide for these speakers, again two explanations are possible: bilingualism or systemic variation.¹ Probably both factors are responsible.

¹ The lexical incidence of η - and \emptyset - is not as pan-dialectal as that of n - and l -. The theoretical η - and \emptyset - in Cantonese are correlated to different extents with the η - and \emptyset - in other Yuè dialects in terms of lexical incidence. η - does not exist in present-day Mandarin, and the theoretical Cantonese η - cannot be identified with any single onset in present-day Mandarin.

9.2.3 [\pm merger of -ŋ and -n]

For certain speakers, -ŋ/k is realized as -n/t after certain vowels. The variation has been reported in Bauer 1979, Yeung 1980 and Ráo et al 1981:274. The set of vowels involved differs in the three accounts. I postpone the discussion of the specific environments for this variation to Section 9.3.1. For the time being, what matters is my observation that any vowel can constitute the environment for the variation, though with different likelihood. In the extreme cases, therefore, the coda -ŋ/k may theoretically be dispensed with. There is, however, no documentation of any individual who has completed the change from -ŋ/k to -n/t, i.e. replacing the former by the latter after every vowel and for every lexical item involved. Thus the merger of -ŋ and -n remains more theoretical than actual.

9.2.4 [\pm merger of T3 and T5]

Although in Section 3.1.2 I have dismissed as unfounded Killingley's suggestion that there are five rather than six tones in mainstream Cantonese, I believe she has faithfully reported the situation in Malayan Cantonese, where only five tones are contrasted. Compared with mainstream Cantonese, the saving of one tone clearly results from a merger of T3 and T5. The transcriptions and descriptions given in Killingley 1985a and 1985b show that the merged tone is realized as a highish even tone (at least in isolation) and is thus more similar to mainstream T3 than T5. As such the merged tone can be regarded as a T3, while T5 is dispensed with.

9.2.4 [\pm split of T1]

The suggested split of T1 has been discussed at length in Section 3.1.3. Although I have adopted a [-split] position in my description, I have also recognized that [+split] is a possible strategy for speakers of a certain age group to cope with the upswing of the HE version of T1. For these speakers, represented by Zōng (1964), Y Cheung (1969) and Yú (1979), there are seven rather than six tones: T1 has split into HE and HF.

9.3 Combinational variations

Combinational variations concern the combination of V with -ŋ, ŋ-/θ- with tone, gw/kw- with V, sibilant onsets with V and æ:/iɔ with coda, together with the variation [ɱ]~[ŋ], which is also treated as a combinational variation owing to the phonological status of the syllabic nasals. These will be discussed one by one in this section.

9.3.1 Vowel with -ŋ

The pronunciation of Bauer's (1979) informant exhibits the following departure from SC (for "standard Cantonese", a term used by Bauer):¹

[4]	<u>SC</u>	<u>REALIZATION</u>
	a:ŋ	always a:n
	ɐŋ	always ɐn
	ɛ:ŋ	occasionally ɛ:n ²
	ɔ:ŋ	occasionally ɔ:n
	æ:ŋ	occasionally æ:n
	eŋ	always eŋ
	oŋ	always oŋ

The pattern of variation shown in [4] is typical of the pronunciation of some young people in present-day Hong Kong. This particular style of pronunciation can be captured by two realization rules:

¹ Occluded versions included.

² Bauer actually says that 'the SC /-ɛŋ/ final was almost always realized as /-ɛn/.'. His detailed data, however, show that this is not the case. What leads Bauer to think that way is that he attributes the non-coronalization of the -ŋ in certain instances of ɛ:ŋ to the item's alternation between ɛ:ŋ and eŋ. As I have argued in Section 4.2.2, several of the assumed ɛ:ŋ alternants should be deemed to have lexicalized, and the non-coronalization of -ŋ in them should count as the non-coronalization of ɛ:ŋ, not of eŋ or the alternating ɛ:ŋ/eŋ. This adjusted picture can be more readily captured in terms of natural classes of vowels than the picture as described originally by Bauer, as can be seen in [6].

$$[5] \quad \begin{bmatrix} -\text{cont} \\ -\text{lab} \end{bmatrix} \rightarrow [+cor] / \begin{bmatrix} V \\ +\text{low} \end{bmatrix} \underline{\hspace{1cm}}$$

$$[6] \text{ Optional: } \eta \rightarrow [+cor] / \begin{bmatrix} V \\ -\text{low} \\ +\text{2nd M} \end{bmatrix} \underline{\hspace{1cm}}$$

[5] says that both -n and -ŋ are realized as -n after low vowels, and [6] that -ŋ is optionally realized as -n after ε:, æ: and ɔ:.

Yeung (1980) also reports the variation -ŋ ~ -n. Her study involves 50 informants, stratified by sex and age. It is strange that in her survey only eŋ and ɐŋ exhibit the variation in question. As far as my own observation can tell, Bauer's single-informant account is a better reflection of the facts of the language than Yeung's account. At any rate the tendency to variation is much higher after a: than after e. Ráo et al's (1981:274) allusion to the variation also goes against Yeung's account. Out of the four examples given by Ráo et al, two are with ɐ and the other two with a:, but none with e.

The significance of the variation is that the resultant non-occurrence of -ŋ after certain vowels gives rise to new V-Cd combinational restrictions. Bauer's informant, for instance, seems to have internalized the following constraint:

$$[7] \quad * \begin{bmatrix} V \\ +\text{low} \end{bmatrix} \eta$$

If the environment for the variation spreads upwards in the vowel space, a constraint with a higher degree of generality will be needed. Variation of ε:ŋ, æ:ŋ and ɔ:ŋ is optional for Bauer's informant. Variation of eŋ has been reported by Yeung. I also observe the realization ɔŋ as en. No vowel, therefore, is completely exempt from constituting the environment for the variation. However, vowel height seems to be directly proportional to the rate of retention of -ŋ.

9.3.2 ŋ/θ- with tone

Cases (ii) (iii) (iv) in [3a] above give rise to the following regularity: subject to the lower-level optional modifications [8], T1-3

entail $\emptyset-$ and T4-6 entail $\eta-$.

- [8] a. $\emptyset- \rightarrow \eta-$ <for case (iii)>
 b. $\eta- \rightarrow \emptyset-$ <for case (iv)>

Depending on whether we treat $\emptyset-$ and $\eta-$ as one onset or two, we can represent the entailment in terms either of a realizational rule as in [9a], or of constraints as in [9b].

$$[9] \text{ a. } \emptyset/\eta \rightarrow \begin{cases} \emptyset- / \overline{[T1-3]} \\ \eta- / \overline{[T4-6]} \end{cases}$$

$$\text{b. (i) } * \begin{bmatrix} T1-3 \\ \eta- \end{bmatrix}$$

$$\text{(ii) } * \begin{bmatrix} T4-6 \\ \emptyset- \end{bmatrix}$$

9.3.3 gw/kw- with vowel

The realization of gw- and kw- is variable before rounded vowels: [g/k] is gaining ground at the expense of [gw/kw]. The following table shows the chronology of different accounts of the environment for the variation, in terms of lexical items or syllables.

[10]	Chén & Bái 1958:7	The item kwɔ:ŋ ⁴ "crazy"
	S Cheung 1972:3	$\begin{Bmatrix} gw \\ kw \end{Bmatrix} \text{ɔ:}\eta$
	Fung 1974	kwɔ:ŋ ([-occl] only)
	Bauer 1979, 1982, 1983	$\begin{Bmatrix} gw \\ kw \end{Bmatrix} \text{ɔ:}(\eta)$
	Yeung 1980:9	gwɔ:(ŋ) ¹
	Luke 1983	$\begin{Bmatrix} gw \\ kw \end{Bmatrix} \begin{Bmatrix} u:n \\ \text{ɔ:}\eta \end{Bmatrix}$

¹ It is not clear if Yeung recognizes the variation kw~ k-. Her survey includes gw/g- items only, not kw/k- items. Yet she cites Yú Bǐngzhāo (*An analysis of Cantonese sounds*, p. 13, date and publisher not given) for saying that 'there is confusion between g and gw and between k and kw'. (Yeung 1980:2).

All of these descriptions except Bauer's are empirically inadequate in that they miss certain forms which are clearly the loci of the variation in question. While Bauer's account covers all the conspicuous forms susceptible to the variation, we have argued in Section 4.2.1.2 that Luke's extension of the environment to u:n is both phonologically and phonetically valid. But u:n is by no means the only rime involving the vowel u: — the rimes u: and u:j must also be included. The full environment for the variation gw/kw- ~ g/k- should then be as follows:

$$[11] \quad \left\{ \begin{array}{l} u: (\left\{ \begin{array}{l} j \\ n \end{array} \right\}) \\ \text{ɔ:} (\eta) \end{array} \right\}$$

[11] is capable of a series of further generalizations. First, the coronalized version of ɔ:η, i.e. ɔ:n, is also a valid environment, e.g.

$\left\{ \begin{array}{l} gw \\ g \end{array} \right\} \text{ɔ:n}^2 \text{dzew}^1$ "Guǎngzhōu". [11] can therefore be adjusted to [12]:

$$[12] \quad \left\{ \begin{array}{l} u: (\left\{ \begin{array}{l} j \\ n \end{array} \right\}) \\ \text{ɔ:} (\left\{ \begin{array}{l} n \\ \eta \end{array} \right\}) \end{array} \right\}$$

Second, -j and -n are the only possible codas after u:, and -η and -n

the only occurring codas after $\left\{ \begin{array}{l} gw \\ kw \end{array} \right\} \text{ɔ:}$. [12] can therefore be further generalized as [13]:

$$[13] \quad \left\{ \begin{array}{l} u: \\ \text{ɔ:} \end{array} \right\} (\text{Cd})$$

Third, by convention, the "(Cd)" can be omitted in a linear representation, resulting in [14]:

$$[14] \quad \left\{ \begin{array}{l} u: \\ \text{ɔ:} \end{array} \right\}$$

Fourth, thanks to the constraints against the combination of gw/kw- with y:, y:n, æ:n, æ:η, ej, en, ow, om and the non-occurrence of gw/kw- with æ: or oη, the following generalized rule can be formulated:

[15] Optional: $\begin{bmatrix} gw \\ kw \end{bmatrix} \rightarrow \begin{bmatrix} g \\ k \end{bmatrix} / \text{---} \begin{matrix} v \\ [+round] \end{matrix}$

For the individuals who adhere to the realization as g- and k- in this environment, the following constraint has in fact been internalized:

[16] * $\begin{bmatrix} gw \\ kw \end{bmatrix} \begin{matrix} v \\ [+round] \end{matrix}$

[16] can also be regarded as a constraint in the making as far as the entire speech community is concerned.¹

9.3.4 Sibilant onsets with vowel

The realization of sibilants discussed at length in Section 8.6 can be formulated as [17] and [18] together:

[17] $\begin{bmatrix} -ant \\ +cor \\ \{-voic\} \\ \{-cont\} \end{bmatrix} \rightarrow \begin{cases} (\text{palatoalveolar} / \text{---} \begin{matrix} v \\ [+round] \end{matrix}) \\ \text{alveolar} \\ \text{alveolopalatal} \end{cases}$

[18] Optional: $\text{alveolar} \rightarrow \text{palatoalveolar} / \text{---} \begin{matrix} v \\ [+high] \\ [-back] \end{matrix}$

[17] says that sibilants are realized as either alveolars or alveolopalatals, and are optionally realized as palatoalveolars when followed by rounded vowels. [18] says that if sibilants are realized as alveolars in general, they are optionally palatalized when followed by high front vowels (in which case they may not be distinguishable from alveolopalatals).

9.3.5 æ:/iɔ with coda

Killingley (1982:13, 1985:26) reports an allophone of æ: in Malayan Cantonese, namely iɔ, which is the obligatory realization of æ: before -ŋ. The variation can be formulated as follows:

¹ See Section 9.5.2 below.

[19] $\alpha: \rightarrow [\text{ɔ}] / _\eta$

9.3.6 $[\eta] \sim [\text{ɱ}]$

Morphemes that belong to the lexical class for which $[\eta]$ is expected are sometimes pronounced $[\text{ɱ}]$ by certain people. The variation is alluded to in Bauer 1979 and treated at length in Bauer 1982, 1983 and 1986. Ráo et al (1981:294) also mention this variation, but regard it as an isolated case of progressive assimilation which applies to the lexical item $[\eta]^5$ "five" only.

Since $[\eta]$ is regarded as underlyingly $\eta\alpha:$ in this thesis, the variation in question can be formulated as follows:

[20] Optional: $\eta\alpha: \rightarrow [\text{ɱ}]$

According to Bauer 1982, some people never use $[\eta]$. They can be considered to have internalized the following constraint:

[21] $\ast \eta \begin{bmatrix} \text{V} \\ +\text{high} \\ +\text{round} \end{bmatrix}$

[21] can also be regarded as a constraint in the making as far as the entire speech community is concerned.

9.4 Lexical-incidental variations

In this section we first discuss the relationship between rules and the lexicon in pronunciation variation. Then we deal with actual cases of lexical-incidental variation. Since by definition such variation can hardly be captured by generalized rules, we shall be content with furnishing examples only.

9.4.1 Rules and the lexicon in pronunciation variation

So far we have used two opposing types of phonological rules: obligatory vs optional. Another distinction can be made of rule types: categorical vs variable. Wells (1982:64) writes:

A categorical rule applies independently of stylistic or social considerations. (...) A variable rule does depend on non-linguistic considerations of this kind.

Harris (1986) distinguishes between phonological variations that are "lexically selective" and those that are "lexically blind".¹ The distinction has to do with the notion "lexical diffusion" in the theory of sound change, first proposed by Wang (1969). Wells (1982:65) explains:

[R]ules sometimes apply only to certain parts of the vocabulary, or more readily to some parts than to others. (...) It appears that sound changes often spread gradually through the vocabulary (...).

As Wells notes, it is not always easy to determine whether a given rule is optional or variable:

Our answer may vary according to whether we are attempting to describe an idiolect (...) or the speech of a whole community. Variability observable within a community does not necessarily imply variability within the individual idiolects that constitute the speech community. (p.65)

By the same token, optional and/or variable rules observable within a community may be idiolectally speaking better characterized as involving variations that are lexically selective, i.e. the locus of variability may not be the phonological rule component but the lexicon. Granted a lexical dimension in pronunciation variation, any such variation above the level of consciousness is theoretically capable of being implemented in the lexicon. A stronger claim than this has been put forward by Harris (1985:302):

[I]nnovations involving phonetically discrete alternations of the type under discussion here [i.e. sociolinguistically constrained alternations between phonemes] are not phonological changes at all but merely changes in the lexical incidence of phonemic units (except insofar as they may eventually lead to phonological restructuring).

¹ He recognizes a third kind, that which is morphologically sensitive. I ignore this particular kind of phonological variation here since it does not bear directly on the exposition in this chapter.

The extent to which the lexicon is involved in "phonetically discrete alternations" is a highly technical issue and beyond the scope of this thesis. What is really relevant at this point is that the majority of variations we have dealt with so far in this chapter involve "phonetically discrete alternations" and exhibit at least some degree of lexical selectivity. There are lexical items that waver between the alternative pronunciations even among the speakers who do contrast the alternants in general. [22] gives two such examples for each of the variations susceptible to lexical selectivity.¹

<u>[22] VARIATION</u>	<u>EXAMPLES</u>
[± merger of n- and l-]	{ ₁ ⁿ }eŋ ^l "carry", { ₁ ⁿ }eŋ ^l "grain"
[± merger of T3 and T5]	si: ^{3/5} "exam", sy:n ^{3/5} "garlic"
[± split of T1]	sa:m ^{HE/HF} "three", gəm ^{HE/HF} "gold"
ŋ- ~ Ø-	{ _Ø ^ŋ }eŋ ^l "hook", { _Ø ^ŋ }a:m ^l "correct"
-ŋ ~ -n	se{ _n ^ŋ } ³ "blow the nose", dz{ _{ok} ^{et} } ^l "at last"
gw/kw- ~ g/k-	k(w)ɔ:ŋ ⁴ "crazy", g(w)ɔ: ³ "to cross"
[ŋ] ~ [m]	{ _[m] ^[ŋ] } ⁵ "five", {[_[m] ^[ŋ]]} ⁴ (a surname)

9.4.2 Semi-regular lexical-incidental variations

Because of the existence of etymologically or synchronic-morphophonologically related lexical items that recurrently involve a certain pair of phonological entities, we can speak of lexically selective morphophonological alternations. As far as pronunciation is concerned, the items involved exhibit lexical-incidental variation. The following table shows the semi-regular alternations involved with corresponding examples of items that have the alternative pronunciations.

¹ Examples are not given for the variation concerning æ: and the sibilant onsets. The alternation [æ:] ~ [iɔ] in Malayan Cantonese seems to be strictly allophonic, in the sense that rule [19] is obligatory and categorical as far as Malayan Cantonese is concerned. The different realizations of the sibilant onsets, on the other hand, may not involve 'phonetically discrete alternations' at all, i.e. the variation may be below the level of consciousness.

[23]	<u>ALTERNANTS</u>	<u>LEXICAL ITEMS</u>
a)	a:/ə	h{ _{a:} ^ə }p ² "box", s{ _{a:} ^ə }ŋ ¹ "unfamiliar"
b)	ɛ:ŋ/en	d{ _{ɛ:} ^ə }ŋ ² "peak", ts{ _{ɛ:} ^ə }k ³ "red"
c)	T3/T2	pow ^{3/2} "shop", dəj ^{3/2} "doublet"
d)	T4/T2	ŋən ^{4/2} "silver", fɔ:ŋ ^{4/2} "room"
e)	T5/T2	la:m ^{5/2} "embrace", nəj ^{5/2} "girl"
f)	T6/T2	wu:j ^{6/2} "meeting", dəw ^{6/2} "beans"
g)	[-voic]+T5/[-cont]+T6	{ty:n ⁵ } "broken", {tsoŋ ⁵ } "heavy"
h)	[-voic]+T4/[-cont]+T6	{ti:w ⁴ } tsa:4 "investigate"
		{dzi:w ⁶ } tsy:n ⁴ ke:j ⁴ "legend"

Hashimoto (1972:116) characterizes (a)-(g) as literary vs colloquial pairs. She actually formulates one rule to provide for (a) and (b) (p.171) and another for (c)-(f) (p.180). Since for her the a:/ə alternation is confined to a:ŋ/en only, cases with -m/p are not provided for. (b) has been discussed in Section 4.3.2 and (c)-(f) in Section 3.2.1. Because of the continuing lexicalization of items with ɛ:ŋ and en and the T2-switched items, the alternations are highly lexically selective nowadays. The same is also true of case (a). As for (g) and (h), they are remnants of (1) processes of derivation by tone change (cf. Kam 1977) that are no longer productive and (2) the historical sound change ONSET ASPIRATION shown in [29], section 7.2.2.1.2. The non-productiveness of all these alternations and their high lexical selectivity render the variations lexical-incidental, though semi-regular.

9.4.3 Merger-suggestive lexical-incidental variations

Another pair of phonological entities recurrently involved in lexical-incidental variations are T3 and T5. Consider the following items that have alternative T3 and T5 versions.

[24]	kəw	"mortar" ¹
	si:	"exam" ²

¹ The pronunciation in T3 is recognized in Hashimoto 1972:332.

² The pronunciation in T5 is recognized in Vance 1976:389.

tsy:	"place"
sɛ:	"house"
sy:n	"garlic"
tow	"rabbit"
ji:(mɔj ⁵)	"barley"
(wu:j ⁶)ji:	"meeting"

Unlike the examples in [23], the alternation shown in these items has no etymological or morphophonological motivation.¹ I suspect that they might be the sign of the very beginning of merger of T3 and T5 proceeding by lexical diffusion. These items take the lead and their alternation constitutes the initial "lag phase" of the diffusion process. The merger of these two tones in Malayan Cantonese is a good indication of the plausibility of their merger in mainstream Cantonese. Moreover the fact that T3 and T5 are the only pair of tones that recur as unmotivated alternants is otherwise inexplicable.

9.4.4 Isolated lexical-incidental variations

Isolated lexical-incidental variations are very numerous. [25] provides five examples with conditioning factors stipulated.

[25]

"eaves":	ji:m ⁴ (prescribed), si:m ⁴ (popular), jɛm ⁴ (colloquial)
"come":	l $\left\{\begin{smallmatrix} e \\ ɐ \end{smallmatrix}\right\}$ j ⁴ (free variation)
"easy"	ji: ⁶ (mainstream), ji: ⁴ (Malayan) (Killingley 1985a)
"advice"	dzoŋl $\left\{\begin{smallmatrix} gok^1 \\ gow^3 \end{smallmatrix}\right\}$ (Kong Kong), (Guǎngzhōu) (Ráo et al 1983:198)
"this"	$\left\{\begin{smallmatrix} li:1 \\ di:1 \end{smallmatrix}\right\}$ (current) (obsolete) (used in Jones and Woo 1912)

9.5 Correlating pronunciation variables with non-linguistic variables

In this section the pronunciation variables we have established so far are correlated with non-linguistic variables, including regional,

¹ Except for 'barley' where the T3 and T5 versions are theoretically represented by two different characters. The etymological difference, however, is no longer transparent and the two characters are used by and large interchangeably nowadays.

chronological, social, and stylistic variables.

9.5.1 Regional variables

Cantonese is spoken as a native language by several million people in Hong Kong and Guǎngzhōu, as well as hundreds of thousands in Macao, Malaya and Singapore. Because of the unquestionable status of Cantonese as the everyday language in Hong Kong, Guǎngzhōu and Macao and the physical proximity and relatively high level of interaction between the three places, we have isolated the variety of Cantonese characteristic of these three places collectively as mainstream Cantonese. Despite the extensive common ground, differences between the Hong Kong and Guǎngzhōu varieties have been suggested in the literature. We can thus speak of "Hongkongisms" and "Guangzhouisms". For want of published data, the regions actually covered in this brief survey are confined to Hong Kong, Guǎngzhōu and Malaya.

We begin with Malayan Cantonese. For information about Malayan Cantonese we rely on a single writer, namely Killingley, who calls herself "a fourth generation Malayan Chinese". It is her "language from birth": she acquired it at home (Killingley 1982b:15-7). The departures of Malayan Cantonese from the mainstream variety are summarized below:

- [26] SYSTEMIC: [+ merger of T3 and T5]
 COMBINATIONAL: æ: → [iɔ] / __ŋ
 REALIZATIONAL: a) T1 → HE
 b) T4 → LLE
 c) T3/T5 → T3
 d) The tone shapes seem to be less intrinsic than those in mainstream Cantonese: the actual realization, especially with respect to orientation (as opposed to register), is very much context-dependent. (1985b)
 LEXICAL-INCIDENTIAL: (examples only)
 ji:⁴ "easy" (T4 mainstream)
 nej⁴ "nun" (1985b:27) (with ej mainstream)

We now turn to Hongkongisms and Guangzhouisms.

Under the assumption that ŋ- and Ø- are contrastive, Yuán et al

mention that recently the majority of Cantonese speakers pronounce \emptyset - as η - while a minority drop η -. In a description that basically does not recognize a contrast between η - and \emptyset - (e.g. the present thesis), Yuán et al's description could be reworded as "the unitary η/\emptyset - has a higher tendency to be realized as η - than as \emptyset -". Ráo et al (1981:270) describe the phenomenon as a "general trend". They actually give almost all the theoretical \emptyset - items the pronunciation with η - in the dictionary entries. In Ráo et al 1983, however, the same items are given alternative pronunciations in η - and \emptyset -. In Hong Kong, the pattern of occurrence of η - and \emptyset - is not skewed in favour of η -. As a matter of fact, among youngsters in Hong Kong, the tendency is just the other way round. (Ráo et al 1981:270)

Another difference between Hongkongisms and Guangzhouisms concerns the realization of T1. According to Zēng (1982:10), Hong Kong tends to use HE while Guǎngzhōu tends to use HF, though the two forms are largely interchangeable in both varieties.

The variation $-\eta \sim -n$ is very probably a Hongkongism. It has never been mentioned in accounts of Guǎngzhōu Cantonese, and Ráo et al (1981:274) expressly characterize it as a Hongkongism.

For lexical-incidental differences between the two varieties, we cite the following.

[27] ITEM	HONG KONG	GUǎNGZHŌU
"advice"	dzon ¹ gok ¹	dzon ¹ gow ³
"seize"	dza:w ²	dza: ¹ (Ráo et al 1983:137)

The Guǎngzhōu pronunciations in [27] are innovations in the development of Cantonese. The changes are clearly due to the influence of Pǔtōnghuà, the codified language variety on the Chinese mainland. Certain lexical items in Pǔtōnghuà have been assigned a pronunciation which is different from the one projected from the Middle Chinese lexical representation. Hence, table [28].

[28] ITEM	PROJECTION from MC	PǔTONGHUÀ
"advice"	zhōnggù	zhōnggào
"seize"	zhāo	zhuā

The Hong Kong pronunciations in [27] correspond to the Mandarin projections of MC (with tone modified in the case of "seize"), while the Guǎngzhōu pronunciations correspond to the codified Pǔtōnghuà pronunciations.

9.5.2 Chronological variables

For the correlation between pronunciation variables and time, we adopt the Labovian apparent-time methodology which compares the speech of older and younger speakers and assumes that differences between them reflect changes in real time.

Two variables can be fairly confidently identified as reflecting sound changes in progress. These are gw/kw- ~ g/k- and [ŋ] ~ [ŋ̚]. The following table, adapted from Yeung 1980:30, shows the % realization with g- for the item gwɔ:k³ "nation".

[29]	AGE	MALE	FEMALE
	12-26	96	97
	20-27	78	83
	30-36	56	67
	40-46	42	58
	50-57	13	33

Bauer's (1982) statistics for the same variation do not reflect a trend as neat as Yeung's do, but the rough tendency that the score varies inversely with age can still be sensed.

Table [30], reproduced from Bauer 1982:106, shows the % [ŋ̚] scores of the pronunciation of the items expected to be [ŋ].

[30]	SPEECH CONTEXT		A G E G R O U P S							
	A		B		C		D			
	(15-22)		(23-30)		(31-44)		(45+)			
	M	F	M	F	M	F	M	F		
	Spontaneous Speech		91	68	61	56	14	69	0	1
	Story		89	72	47	50	0	67	2	0
	Word Lists		77	30	21	21	0	66	3	0

Except for the unexpectedly high scores for C/F, the scores indicate a general correlation between age and score: the younger the person, the higher the score.

To the extent that n- and l- were surely contrastive in Cantonese some time in the past, the present-day merger status recognized in this thesis represents an innovation. However, it does not follow that the merger has taken place recently enough to be reflected in the speech of different age-groups in the present-day population. Table [31], adapted from Yeung 1980:25, shows the % [l-] score for the items with a theoretical n-.

[31]	AGE	M	F
	12-16	100	100
	20-27	100	95
	30-36	100	88
	40-46	96	73
	50-57	94	65

The figures in [31] must be interpreted with great care. For speakers that do not contrast n- and l-, whom I hold to be the overwhelming majority, the onset l- (incorporating the theoretical n-) is realized not only as [l], which is usual, but also occasionally as [n]. It follows that realizing a theoretical n- as [n] does not necessarily mean that the speaker maintains a contrast of n- vs l-, or is conscious of the [n] vs [l] distinction at all. Moreover, I have suggested earlier that the n- vs l- contrast maintained by some speakers may be due to analogy to other Chinese dialects. The structure of the population of Hong Kong with respect to language is such that the older a person is, the less likely he is to be a "pure" native speaker of Cantonese, i.e. the more likely it is that he also speaks another Chinese dialect, usually his parents' tongue. The bilingualism hypothesis is corroborated by the considerably wider discrepancy between male and female scores than that for other variations studied by Yeung. While intrinsic difference in male vs female variation patterns cannot account for the wider discrepancy, it can be explained in terms of bilingualism in accordance with another hypothesis: that among the first- and second-generation

immigrants there is a higher degree of integration with the Hong Kong society in males than in females.

Two other variations can also be shown to be correlated with time. [32], adapted from Yeung 1980:44, shows the % [-n] score for the standard -ŋ items.

[32]	AGE	M	F
	12-16	11	9
	20-27	3	1
	30-36	0	0
	40-46	0	0
	50-57	0	0

While the scores show that the realization of -ŋ as [n] is a very recent development, the effect of the variation is somewhat underrepresented because Yeung has not confined the environment to after e and ə, the only environment she observes to be possible for the variation.

Another variation correlated with time is one between the [+fall] and [-fall] versions of T1, i.e. between HF and HE. We have seen in Section 3.2.2 that the default value of T1 has changed from HF in the course of the present century. The variation, however, has not been dealt with in any sociolinguistic study.

The variation between ŋ- and Ø- has also been suggested to represent real-time change. Yeung's (1980:34-8) survey of the variation (treated as one between [ŋ]- and [ʔ]-) has the following concluding remarks:

There are 601 incidences (according to Wong's [1940] syllabary) of [ŋ-] words in the utterances of the 50 speakers. Of these, 480 are actually pronounced with [ŋ-]. (...) for [ʔ-], the figures are 1338/1760, or 76%. Wong's syllabary was compiled in 1938 and based on the pronunciation in Canton. When applied to the present day Hong Kong, for [ŋ-] words, it is correct for 4/5 of the times, and for [ʔ-], 3/4 of the times. These figures are not too high considering that only 42 years have passed.

Yeung's is an oversimplified account of a highly complex situation.¹ First, the % "correct" scores are not 100% even for the age-group 50-57: 92% (110/119) for the theoretical η - and 71% for the theoretical \emptyset -. Second, as we have seen in Section 9.2.2, realization of the theoretical \emptyset - as η - was already predominant in Guǎngzhōu in 1947. Third, the variation is one over three classes of morpho-syllables as shown in [3], Section 9.2.2 above, rather than just over two classes (her [η -] items and [\emptyset -] items) as she thinks. Fourth, by assuming a contrast between η - and \emptyset -, she views the variation as one of lexical-incidence only, but as we have seen it actually involves systemic and/or combinational differences. For these reasons, we must conclude that until surveys more meticulously designed are carried out, little can be said about how the variation under consideration is related to time, or indeed to any non-linguistic variable.

9.5.3 Social variables

The social dimension of pronunciation variation can be conceived of broadly or narrowly. In the broad sense, it refers both to the cross-classifying characteristics of speakers such as sex, educational level, income, profession, age, etc. and to different situations of discourse. In the narrow sense of the term, on the other hand, it excludes age, which is dealt with separately as the shadow of the time dimension, and the situation of discourse, which is dealt with separately as representing stylistic differences. Here we take the narrow sense.

Among the potential social variables, only sex and education level have been related to pronunciation variables in Yeung 1980 and Bauer 1982. The variation [\ast palatoalveolar] for sibilants and the variation $-m \sim -n$ as studied by Yeung do not give results that are significant enough to warrant a report in this thesis, and the actual figures show that even if the variations can be established, sex difference is not significant for either variation. Apart from these two variables, [32] shows that sex difference is not significant for the $-\eta \sim -n$ variation either. [29], on the other hand, shows that the pronunciation of the female is regularly more advanced than that of the male for the

¹ Cf. Section 9.2.2 above.

variation gw/kw- ~ g/k-. The n- ~ l- variation shows considerable sex difference, as can be seen in [31], but the difference is capable of different interpretations, as discussed in the last section (9.5.2). Yeung's figures with respect to the [ŋ-] ~ [ʔ-] variation also show considerable sex difference. However, for the same reasons as given in Section 9.5.2, precise inference from the figures is impossible.

Bauer's figures with respect to the sex difference for the gw/kw- ~ g/k- variation do not agree with Yeung's. His figures show that the males are more advanced than the females for most of the age groups. He says that at the time he has "no plausible explanation to account for why these differences exist"(p.212). As for the variation [ŋ] ~ [m̩], as [30] shows, the age group C (=31-44) exhibits a dramatic difference between the two sexes, and group A/F has a higher tendency than A/M to shift back to the standard, conservative pronunciation.

Thus, as far as sex difference is concerned, only the gw/kw- ~ g/k- variation, as studied by Yeung, shows significant and regular sex difference.

Education level is considered in Bauer 1982 but not in Yeung 1980. For both the [ŋ] ~ [m̩] and gw/kw- ~ g/k- variations studied by Bauer, I cannot extract any pattern of variation with respect to education level.

9.5.4 Stylistic variables

Register differences are considered in Bauer 1982 but not in Yeung 1980. Thus we have relevant figures for the [ŋ] ~ [m̩] and gw/kw- ~ g/k- variations only. For the first variation [30] shows that in general the more formal the style, the higher the tendency to conform to the standard, conservative pronunciation. The figures for the second variation do not reflect as neat a pattern, but a rough trend along the same line still holds.

Apart from those variations, the stylistic overtone of the variation HE ~ HF in present-day Cantonese has been discussed in Section 3.2.2.

CHAPTER 10: BEYOND THE SYLLABLE

Our discussion has so far been confined to the syllable. Syllable-bound description is in fact standard practice in general accounts of the sounds of Cantonese, and indeed of any Chinese dialect. The typical defence for such practice is that "most of the morphemes in Chinese are monosyllabic" (Hashimoto 1972:87). However, since the subjective inventory of morphemes is highly variable from person to person, while the inventory of well-formed syllables is relatively fixed, a more meaningful characterization of the lexicon-syllabary relation is that every monosyllable could, when pressed, be associated with some "meaning", either lexical or functional, and is therefore a potential "word". It is this high lexical potential of the monosyllable that leads linguists to analyse each monosyllable as an isolatable (if not categorically "isolated") citation form. Thus, syllable boundaries coincide with the psychological boundaries of the maximally isolated lexemes in citation form.

Kao (1971:13) is probably right in saying that "before any investigation of Cantonese running speech may be undertaken, it is of primary importance to arrive at a systematic characterization of the phonological features of the syllable." The preeminence of the syllable, however, should not be an excuse for not going beyond the syllable. Thus the present chapter goes beyond the syllable and considers connected speech. Our discussion divides into two main parts, one dealing with casual speech (CS), which presupposes connected speech (as opposed to isolated citation forms), and the other with intonation.

10.1 Casual speech

The phonology of the Cantonese monosyllable is basically a phonology of the maximally isolated lexemes in citation form. In contrast to this, phenomena of casual connected speech have barely been touched on. As Lass (1984:295) observes, "modifications in connected speech can often have effects with higher-order systematic relevance." In particular, "we must consider not only processes that are simply a function of contiguity in the speech chain, but also ones tied to tempo,

and register and style as well." Phonological regularities in CS should therefore form an integral part of the phonological description of a language. This section (10.1) is devoted to a discussion of CS phenomena.

Since the syllable boundary \$ coincides with the maximally isolated lexeme boundary #, \$ can actually be dispensed with in Cantonese phonology. In my formulation, # is more psychological and potential than intrinsic. As such # is subject to deletion, contingent on a cluster of factors including tempo, formality, whether a lexical item is analysable into morphemes, the information content of the item involved (emphasized, neutral, weakly stressed, etc. at the utterance level), etc. At the present state of knowledge, we know little about the precise conditions for the presence or absence of #. All I could say is that # bars any CS processes from taking place, i.e. CS processes are possible only when and where # is lifted.

There is very little description of Cantonese casual speech; this is hardly surprising in view of the fact that "perception is a function of expectation" (Wang 1969). Lass (1984:296) elaborates on this very point:

[I]t is perfectly possible for the listener to 'hear' things that aren't there, or things quite different from what's actually being said (...)

Thus, unless one is meticulously listening out for CS modifications and is properly trained to do so, such modifications can easily escape one's attention. Lass (1984:298) enumerates the primary CS characteristics:

- (a) increasing frequency of assimilation, i.e. loss of distinction between neighbouring segments, often as a result of
- (b) suppression of boundaries, leading to the reorganization of word-size chunks of phonic material into syntactic groups with multiple membership, treated as single phonological words;
- (c) lenition, especially by opening of stricture;
- (d) vowel reduction (primarily shortening and centralization), with vowel loss leading to syllabification of consonants;
- (e) Shortening of long segments;
- (f) reduction of clusters.

Subject to minor adjustment, all of these characteristics can be found in Cantonese CS. The lifting of # mentioned above, for instance, corresponds to (b) here. For ease of exposition, we divide the

discussion of Cantonese CS into three sub-topics, namely cross-syllable assimilation, inter-syllabic onset lenition, and contraction.

10.1.1 Cross-syllable assimilation

There are two kinds of cross-syllable assimilation, namely contiguous and non-contiguous. These are discussed in the following sections. Contiguous or not, the assimilations to be discussed correspond to CS characteristic (a) above.

10.1.1.1 Contiguous assimilation

Consider the following juxtaposed IC (for "isolated citation") forms and CS forms, taken from Yuán et al 1960:193-4.

[1]	IC	CS	GLOSS
	[m] ⁴ how ²	[m] ⁴ ow ²	don't
	{ts} _k em ⁴ jet ⁶	{ts} _k em ⁴ et ⁶	yesterday
	gem ¹ jet ⁶	gem ¹ et ⁶	today
	[m] ⁴ hej ⁶	[m] ⁴ ej ⁶	be not

They all involve progressive assimilation in place and manner of articulation of h-/j- to m-.¹

The following, from various sources, are regressive assimilations:

[2]	SOURCE	IC	CS	GLOSS
a)	Yuán et al 1960:194	sa:m ¹ mu:n ⁴	-[m]	shut the door
b)	Hashimoto 1972:114	je ⁶ bu:n ²	-[p]	Japan
		se ³ mən ¹ dzej ²	-[m]	child
		ga:k ³ meŋ ⁶	-[p]	revolution
		se ¹ pow ⁵	-[m]	bride

¹ I do not accept Hashimoto's (1972:114) example tsi:⁴ne⁶ deriving from tsi:n⁴jet⁶ 'day before yesterday'. As for Ráo et al's example toŋ⁴now¹ from toŋ⁴how¹, 'crowndaisy chrysanthemum', it is alien to me.

c) Bái 1982:19	sew ¹ mej ¹	-[m]	finally
	dze ¹ k ¹ bi:n ¹	-[p]	by the side
	ke ⁵ w ⁵ mo ⁵	-[m]	maternal uncle's wife
d) Hashimoto 1972:114	to ⁵ ŋɔ: ⁶	-[ŋ]	hungry
e) Bái 1982:19	ta: ³ to: ⁴ j ⁴	-[t]	collapse
	tse: ⁴ dew ²	-[n]	broad bean
f) Hashimoto 1972:114	je ⁶ t ⁶ gwɔ: ¹ ŋ ¹	-[k]	sunlight
	ga: ³ k ³ dzek ¹	-[t]	to sack
g) Yuán et al 1960:294	ba: ⁶ k ⁶ dzok ¹	-[t]	centipede

Of these, (a) to (c) are clearly examples of the (not uncommon in Cantonese) assimilation of a coda to the following bi-labial (i.e. [+ant, -cor, -dist]) onsets in place and/or manner of articulation.¹ (d) and (e) are examples of the less common regressive assimilation to dentals and velars. In particular, the CS forms in (e) appear to be lexicalized for some speakers, as is the CS form for "bride" in (c). (f) and (g) are probably peculiar to those who confuse -ŋ/k with -n/t in certain contexts (Cf. last chapter).

Contiguous assimilation, progressive or regressive, is very sporadic. Before we can better understand the conditions for its application (apart from the meta-condition that assimilation does not take place across #), all we can say is that the processes in question are lexically highly selective.

10.1.1.2 Non-contiguous assimilation

Consider the following juxtaposed IC and CS forms taken from Bái 1982.

¹ Ráo et al (1981:293) also cites *jet¹doŋ⁶mu:n⁴* from *jet¹dow⁶mu:n⁴*. I guess it is *jet¹do⁶mu:n⁴* that they are referring to, which form should be classed with the examples in [14] below. highly selective.

[3]

a)	<i>dʒa:3ga:lji:l</i>	<i>ga:3ga:lji:l</i>	fake
	<i>jy:n5nem4nem4</i>	<i>jy:n5jem4jem4</i>	soft
	<i>tsɛ:ŋ1tɔ:j4</i>	<i>tsɛ:ŋ1tso:j4</i>	moss
b)	<i>mu:n4tsa:ɲ5</i>	<i>mu:n4tsa:n5</i>	threshold
c)	<i>ga:p3sa:ŋ1</i>	<i>ga:ɲ3sa:ŋ1</i>	raw
	<i>ga:p3ŋa:ŋ2</i>	<i>ga:ɲ3ŋa:ŋ2</i>	by force
d)	<i>bət1jɣ:4</i>	<i>bət1ji:4</i>	would be better
	<i>mu:n4sɣ:ldɔw2</i>	<i>mu:n4si:ldɔw2</i>	pivotal socket for door

(a) represents the complete assimilation of an onset to an adjacent (but not contiguous, of course) onset. (b) is the coda counterpart of (a). However, since both the input and output of the change are nasals in (b), we could say that the change involves place of articulation only. (c) is the [*occl] counterpart of (a) and (b). (d) is the assimilation in [*round] to the adjacent V.

The assimilations in (3) are again applied very sporadically and can therefore be characterized as lexically highly selective.

10.1.2 Non-initial onset lenition

Non-initial onset lenition involves voicing, tapping, and stricture opening. It corresponds to Lass' CS characteristic (c).

The phonetic voicing of [+voic, -cont] onsets has been discussed in Section 8.5 when we were dealing with the realization of onsets. I am not repeating myself here; suffice it to say that in the context of the present chapter we can see clearly that the optional phonetic voicing of those onsets in the absence of a contiguous # is a particular case of onset lenition in CS.

The phonetically voiced intervocalic [d]- may be further lenited into a tap, i.e. [ɾ]. It is remarkable that tapping can apply to l- as well.

The tap that results from the lenition of l- is different from [r]: a tapping kind of obstruction is formed centrally, while air passes out at the sides without interruption. The tap so produced can be called a "lateral tap". The symbol for "alveolar lateral flap" is "J" on the 1979 IPA chart. The (optional) process involved can be formulated as follows:

$$[4] \quad \begin{bmatrix} +\text{cor} \\ +\text{cont} \\ +\text{voic} \end{bmatrix} \rightarrow \text{tapped} / [+cont] \text{ ___}$$

The items that often undergo [4] include the following:¹

[5] "these"	li:1	} [r] i:1
"those"	gɔ:2	
"we"	ŋɔ:5	} [r] ej6
"you"	lej5	
"they"	kej5	
"remember"	gej3	[r] ɐk1
[6] "cream"	gej6	[J] i:m1
"chocolate"	dzy:1	gu:1 [J] ek1
(particle) ²		[J] a:k3
(particle)		[J] ɔ:4

The favourable environment for the application of [4] is fast tempo. In particular the reduction of the preceding syllable, which reduction will be discussed in the next section (10.1.3), is highly correlated with the application of [4].

In Chapter 8 we have seen that there may be labial friction in w- and palatal friction in j-. In CS, non-initial w- or j- is typically realized in the more open version, i.e. as a vocoid.

10.1.3 Contraction

Contraction often takes place when two syllables are no longer separated by a #. Some contractions involve the deletion of a non-final

¹ All [r]s would be d- and all [J]s would be l- in IC forms.

² Note that Cantonese particles are enclitics by nature.

coda, some involve the deletion of non-marginal coda and onset. I shall show that contraction phenomena can be best captured by reference to the moraic organization of the syllables involved, when at the same time we shall see that even in cases where no segmental deletion is involved it may still make sense to speak of contraction.

10.1.3.1 Mora deletion

Consider the following description by Jones and Woo (1912:xiii):

e alone is occasionally substituted for ei in unimportant words when pronounced very short; (...)

o alone is occasionally substituted for ou in unimportant words when pronounced very short; (...)

The relation between the shortening of the syllable and the substitution of e and o for ei and ow is obvious. It is strange, however, that while shortening of the syllable of neutral length by half (i.e. from quaver to semiquaver or from two semiquavers to two demisemiquavers) happens to syllables with any rime, deletion of coda applies to ej and ow only. It can be argued that as far as the data in Jones and Woo 1912 are concerned, what actually takes place is the deletion of mora, not of coda. This is true of all three types of rime:

- [7] a. $\begin{array}{c} M \quad M \\ | \backslash / \\ O \quad V \end{array} \rightarrow$ b. $\begin{array}{c} M \\ | \backslash \\ O \quad V \end{array}$
- [8] a. $\begin{array}{c} M \quad M \\ | \backslash / | \\ O \quad V \quad Cd \end{array} \rightarrow$ b. $\begin{array}{c} M \\ | \backslash \\ O \quad V \quad Cd \end{array} \rightarrow$ c. $\begin{array}{c} M \\ / | \backslash \\ O \quad V \quad Cd \end{array}$
- [9] a. $\begin{array}{c} M \quad M \\ | \backslash | \\ O \quad V \quad Cd \end{array} \rightarrow$ b. $\begin{array}{c} M \\ | \backslash \\ O \quad V \quad Cd \end{array} \rightarrow$ c. $\begin{array}{c} M \\ / | \backslash \\ O \quad V \quad Cd \end{array}$

[7] is the simple shortening of the unchecked V, which answers Lass' CS characteristics (d) and (e). Note that [8] and [9] are neutralized as a result of the deletion of a mora. The reduction of syllables with trochaic rimes as in [8], however, is rare, both as recorded in Jones and Woo 1912 and according to my own observation. Neutralization is

thus virtually avoided. The moraic interpretation of syllable reduction allows for shortening of the syllable with the segmental configuration unchanged. [9], however, covers cases with glide-checked rimes, i.e. ej, ow, ej, ej, ew, which are diphthongs phonetically. When the V is high enough, the brevity of the rime may render the diphthong very similar to a monophthong. Hence Jones and Woo's e and o in place of ej and ow.

The change from (a) to (b) in [7], [8] and [9] can be formulated as a single rule:

$$[10] \quad \begin{array}{c} \text{S} \\ / \quad \backslash \\ \text{M} \quad \text{M} \end{array} \rightarrow \begin{array}{c} \text{S} \\ / \quad \backslash \\ \text{M} \quad \text{ } \end{array} / \text{ ___ } X \quad (\text{Condition: } X \neq \#)$$

[10] is a variable rule, as all processes for CS are expected to be. There are conditions that favour or disfavour the application of [10]. In view of the rareness of [8], we can say that the mora with branching association lines is not favoured for deletion.

10.1.3.2 Coda deletion

We said earlier that shortened syllables did not necessarily result in coda deletion. Deletion of coda, however, does occur. It frequently occurs, for instance, when a shortened syllable precedes a lateral. Thus, in addition to [11], which are already provided for by [10] above, there are examples like [12] which clearly result from the deletion of coda.

- [11] a. fi(:)¹li:ls^{a:2} "freezer"
 ha:j¹ka(:)l¹la:ls^{i:2} "high-class"
 fu(:)⁴lok¹ "fluke"
- b. si(:)⁴li:ls^{a:4}la:4 (onomatopoeic item suggesting speed)
- [12] a. pe(j)¹lej¹si:2 "place"
 ge(j)⁶li:m¹ "cream"
 bɐ(t)⁶la:n¹dej² "brandy"
 ke(k)⁶lek¹dzi:2 "clutch"
 be(k)⁶lek¹ "brake"

- b. hɛm⁶ba(:ŋ)⁶la:ŋ⁶ "all"
 dzek⁶bɛ(t)l¹lɛt¹ "straight as a ramrod"
 jɛt¹pa(:j)⁴la:j⁴ "in rows"
 jɛt¹gɛ(w)⁶lɛw⁶ "in pieces"
 gɔ(:k)l¹ɔ:k¹tɛw² "corner"
- c. pɛ(k)l¹lɛk¹pa(:k)l¹la:k¹ "cracker-like sound"
 kɛ(ŋ)l¹lɛŋ¹ka(:ŋ)l¹la:ŋ¹ "clang"
 kɛ(ŋ)l¹lɛŋ¹kwa(:ŋ)l¹la:ŋ¹ "clang"
 fi(:)⁴li:l¹fɛ(:t)⁴lɛ:t⁴ "sobbing sound"
- d. dɛ(k)l¹lɛk¹da(:k)l¹la:k¹ "sound of rain or steps"

The following rule is formulated to provide for the deletion under discussion:

$$\begin{array}{c}
 M \\
 \uparrow \\
 [13] \quad Cd \rightarrow \emptyset / _ l-
 \end{array}$$

[13] says that a demoraed coda before l- is deleted. It is ordered between stages (b) and (c) in [8] and [9], so that it can be stated in this simple shape.

There are also clear cases of the deletion of a demoraed coda other than before an l-. For example:

- [14] tɛɔ(t)l¹mi:n⁶ "outside"
 dze(k)l¹hɛk¹ "immediately"
 dze(k)l¹hɛj⁶ "be the same as"
 dɛɔ(w)⁶hɛj⁶ "be" (emphatic)
 kɛ(j)⁵wa:6 "he says"
 jɛɔ(t)l¹hɛj⁶ "alternatively"
 kɛ(j)⁵[r]ɛj⁶ "they"

Granted that coda deletion is more likely before l- than before other onsets, [14] nevertheless exhibits the same process of coda deletion as [11] and [12]. [13], therefore, can in fact be replaced by a

more general rule (which is variable anyway), one with the environment removed:

M
†
[15] Cd → ∅

Since like [13], [15] is ordered between stages (b) and (c) in [8] and [9], one effect of [15] is the avoidance of [8c] and [9c]. There is good phonological motivation for such avoidance: [8c] and [9c] have an M that is shared by more than two segments, thus violating the hypothesis that half a mora is the threshold of length awareness in Cantonese phonology, which has worked satisfactorily so far. Moreover, as we shall see in Section 10.1.3.4, generalized coda deletion is taken for granted in the process of bisyllabic fusion. As a matter of fact, while half-length syllables, first reported in Jones and Woo 1912, are not uncommon in present-day Cantonese, it is unlikely nowadays for such half-length syllables to have all three segments retained.

10.1.3.3 Cl- formation

The forms in [11] and [12] are just a short step from forming Cl-clusters in CS. The additional step required is the deletion of everything in the reduced syllable other than the onset, including the (now semi-moraic) rime, the only mora left, and the tone. [16] illustrates the kinds of deletion involved:

[16] f(i¹)li:l²sa:2 "freezer"
 g(ɔ³)lɔ:k¹l²ɬw² "corner"
 p(e¹)lek¹p(a¹)la:k¹ "cracker-like sound"

Not all forms in [11] and [12] can be so contracted. Thus the forms in [11b] and [12d] cannot undergo further deletion. As we have seen in Section 5.1.4 when we dealt with Cl- clusters, clustering with l- is confined to labials (including labiodentals which in turn include labiovelars) and velars. To provide for the kind of deletion illustrated in [16], I formulate the following rule:

[17] $M \rightarrow \emptyset$
 |
 $V \rightarrow \emptyset / [-\text{cor}] \text{ ___ } 1-$

Mention of the disappearance of tone is saved by positing principle [18], together with rule [19], which is to be ordered immediately after [17] to prevent the tone (T) from reassociating with an adjacent mora.

[18] The mora is the minimal carrier of tone.

 M
 †
 [19] $T \rightarrow \emptyset$

[18] works both positively and negatively. It explains why shortened, i.e. mono-moraic, syllables have their tone retained in general, and why a tone is not retained when it is not associated with any mora. [18] will be invoked again in the following section (10.1.3.4).

10.1.3.4 Bisyllabic fusion

Cl- formation results in two syllables being realized as one syllable. As such it can be viewed as a process of bisyllabic contraction. The contraction, however, is very much skewed in favour of the second syllable in the sense that the second syllable is completely retained whereas the first syllable is altogether deleted save for the onset, which remains somewhat "extra-syllabic" in the resultant syllable. There is another kind of bisyllabic contraction, where reduction takes place in both syllables. Consider the following examples.

GLOSS	IC	C O N T R A C T E D	
		Plain	Coerced
[20] soon	dzəw ⁶ lej ⁴	dze ⁶ ej ⁴	dzej ⁶⁺⁴
be (emphatic)	dzəw ⁶ hej ⁶	dze ⁶ ej ⁶	dzej ⁶
of course	geŋ ² hej ⁶	ge ² ej ⁶	gej ²⁺⁶
really	dzen ¹ hej ⁶	dze ¹ ej ⁶	dzej ¹⁺⁶
certainly	sət ⁶ hej ⁶	se ⁶ ej ⁶	sej ⁶
alternatively	jət ¹ hej ⁶	jə ¹ ej ⁶	jəj ¹⁺⁶
a moment	jət ¹ dzən ⁶	jə ¹ ɛn ⁶	jən ¹⁺⁶
in the past	gəw ⁶ dzən ⁶ (si:2)	ge ⁶ ɛn ⁶	gen ⁶
radio	səw ¹ jəm ¹ (gej ¹)	sə ¹ ɛm ¹	səm ¹
[21] available	jəw ⁵ dək ¹	jə ⁵ ɛk ¹	ja:k ⁵⁺¹
There's one ...	jəw ⁵ jət ¹	jə ⁵ ɛt ¹	ja:t ⁵⁺¹
all day long	jət ¹ jət ⁶ (dow ³ hək ¹)	jə ¹ ɛt ⁶	ja:t ¹⁺⁶
let it be	jəw ² dək ¹ (kej ⁵)	jə ² ɛk ¹	ja:k ²⁺¹
[22] go out	tset ¹ hej ³	tse ¹ ej ³	tsej ¹⁺³
[23] know	dzi:l ¹ dow ³	dzi ¹ ow ³	dzi:w ¹⁺³
be called	gi:w ³ dzow ⁶	gi ³ ow ³	gi:w ³⁺⁶
where	bi:n ¹ dow ⁶	bi ¹ ow ⁶	bi:w ¹⁺⁶
here	li:l ¹ dow ⁶	li ¹ ow ⁶	li:w ¹⁺⁶
[24] morning	dzi:w ¹ təw ⁴ (dzow ²)	dzi ¹ ɛw ⁴	dzi:w ¹⁺⁴
[25] not yet	mej ⁶ tsəŋ ⁴	me ⁶ ɛŋ ⁴	mɛ:ŋ ⁶⁺⁴
in actual fact	kej ⁴ sət ⁶	ke ⁴ ɛt ⁶	kɛ:t ⁴⁺⁶
persistently	səŋ ⁴ jət ⁶	se ⁴ ɛt ⁶	sɛ:t ⁴⁺⁶
tomorrow	təŋ ¹ jət ⁶	te ¹ ɛt ⁶	tɛ:t ¹⁺⁶
remember	gej ³ dək ¹	ge ³ ɛk ¹	gɛ:k ³⁺¹
ping pong	bəŋ ¹ bəm ¹ (bo:l)	bə ¹ ɛm ¹	bɛ:m ¹

¹ Cf. the contracted forms mi:ɣəŋ⁴ and mi:ɛŋ⁴ cited in Yuán et al 1960:195.

[26] tape recorder	lok ⁶ jəm ¹ (gej ¹)	lo ⁶ əm ¹	lɔ:m ⁶⁺¹
[27] that day	gɔ:2jət ⁶	gɔ ² ət ⁶	gɔ:t ²⁺⁶
[28] only	dzeŋ ⁶ həj ⁶	dze ⁶ əj ⁶	dze:6
[29] especially	jɐw ⁴ kej ⁴ (si:6)	jɐ ⁴ ej ⁴	jɛ:4
[30] go back home	fa:n ¹ ɲok ¹ (kej ²)	fa ¹ ok ¹	fɔ:k ¹
[31] already	ji:5geŋ ¹	ji ⁵ eŋ ¹	jeŋ ⁵⁺¹
self	dzi:6gej ²	dzi ⁶ ej ²	dzej ⁶⁺²
[32] for example	pej ³ ju:4	pe ³ y:4	py:3+4
however	bət ¹ gwɔ:3	bɐ ¹ ɔ:3	bɔ:l+3
now	ji:4ga:1	ji ⁴ a:1	ja:4+1
what	mət ¹ lɛ:5	mɐ ¹ ɛ:5	mɛ:l+5 ₁
it'd be better	bət ¹ lɔj:4	bɐ ¹ y:4	by:l+4
[33] be fond of	dzoŋ ¹ li:3	dzo ¹ i:3	—
therefore	sɔ:2ji:5	sɔ ² i:5	—
[34] ...I guess.	gwu:2a:3	gwu ² a:3	(gwa:2+3) ²
he says	kej ⁵ wa:6	kɐ ⁵ a:6	(kwa:5+6)

We introduced the concept of "coercion" in Section 3.2.1. when we dealt with complex tones. Here we have cases of the coercion of segments. The exact manner of segmental coercion will become clear in a short while. Meantime what we should attend to is the fact that the kind of bisyllabic contraction under discussion may result in two different

¹ Cf. the contracted forms mi:1ɛ:5 and mɛ:5(lengthened) cited in Yuán et al 1960:195.

² Some people use a lexicalized gwa:3 throughout.

contracted CS forms, one plain, the other "coerced". Often it is the coerced form that is more frequently heard. In any case, the coerced form is more easily detected because it is more different from the IC form. By looking at the coerced forms only and ignoring the plain contracted forms, one easily becomes lost in the multiplicity of modifications involved for which generalization can hardly be made. The key to generalization, I claim, lies in the plain contracted forms. By reference to the moraic organization of the syllables concerned, one simple rule is primarily responsible for deriving all the plain contracted forms, which in turn serve as a link between the IC forms and the coerced contracted forms.

In the examples listed above, the rime of the second syllable in the IC form is either iambic, i.e. with short V and long coda ([20] to [31]), or uniform, i.e. unchecked ([32] to [34]). The rime of the first syllable, on the other hand, may be uniform as in [31], iambic as in [20], or trochaic (long V and short coda) as in [24]. Regardless of the moraic status of the two syllables involved, the plain contracted forms can all be primarily accounted for by rule [35], which says that a non-marginal M is deleted.

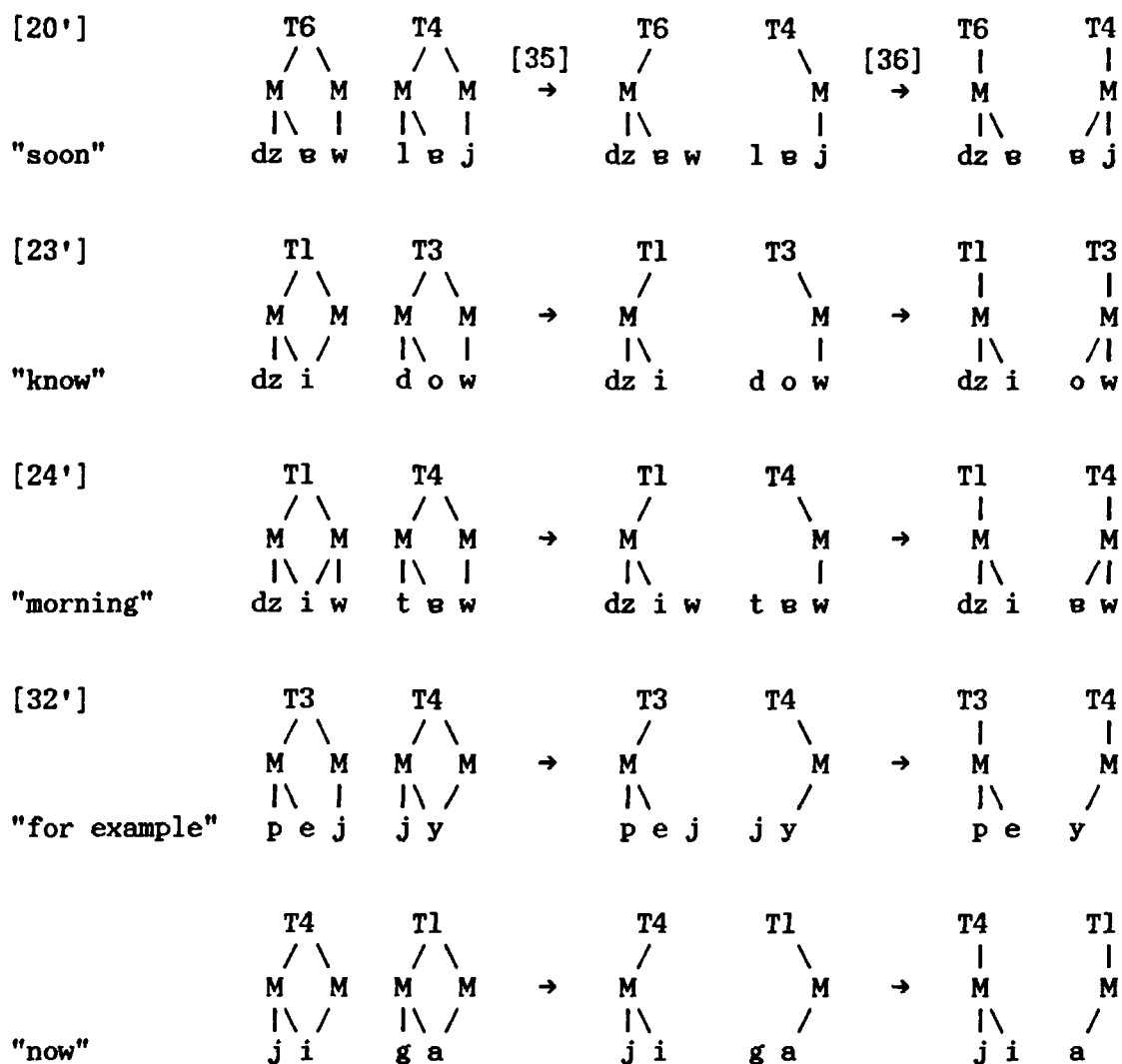
[35] $M \rightarrow \emptyset / X$ <Condition: $X \neq \#$ >

[35] is supplemented by the deletion of the demoraed coda and onset. For this purpose, all we need is not an additional rule but a revision of [15] as [36]:

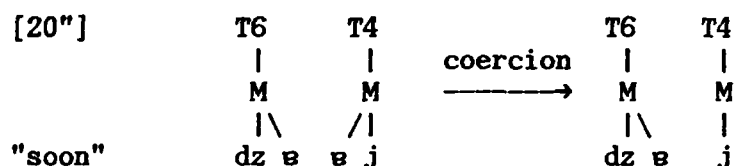
[36]
$$\begin{array}{c} M \\ \dagger \\ \{Cd\} \\ 0 \end{array} \rightarrow \emptyset$$

The contraction of bisyllables in all five different combinations in moraic terms¹ can accordingly be illustrated by the following derivations of concrete examples:

¹ The case of a trochaic rime followed by a uniform rime does not exist in the examples.



The processes outlined above account for all the plain contracted forms. It is quite beside the point to ask if there is one syllable or two in the plain contracted form. One thing, however, is clear: the plain contracted forms are all segmentally different from wellformed syllables. It is probably because of this distance from the wellformed syllables that the plain contracted forms are often "coerced" into segmentally wellformed syllables. For example:



Different segmental configurations in the plain (contracted) forms derive different coerced forms. The coercion in [20], for example, can be generalized as follows:

$$[20*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ ʋ} & \text{ʋ} \text{ Cd} \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & | \\ O \text{ ʋ} & \text{Cd} \end{array}$$

However, as is predictable from [18], i.e. "the mora is the minimal carrier of tone", [20*] runs into trouble when the second syllable is [+occl]: to the extent that the second mora in the coerced form is the sole carrier of the tone of the second syllable in the IC form, it must not be monopolized by a phonetically voiceless segment. Thus an earlier rule [21*] precludes this situation:

$$[21*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ ʋ} & \text{ʋ} \text{ Cd} \\ & [+occl] \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ ʋ} & \text{Cd} \end{array} = \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ a} & \text{Cd} \end{array}$$

[22*] is the e counterpart of [20*]:

$$[22*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ e} & \text{e} \text{ Cd} \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & | \\ O \text{ e} & \text{Cd} \end{array}$$

[23*] is very different: here the two V's are not the same:

$$[23*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ i} & \text{o} \text{ w} \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ i} & \text{w} \end{array}$$

[24] suggests that [23*] can be made more general. Thus:

$$[23/24*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ i} & \text{V} \text{ w} \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ i} & \text{w} \end{array}$$

Then, to move down the examples:

$$[25*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ e} & \text{ʋ} \text{ Cd} \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / | \\ O \text{ ɛ} & \text{Cd} \end{array}$$

[26*] is the [+back] counterpart of [25*].

$$[26*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O & o \quad \mathfrak{e} \quad Cd \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / | \\ O & \mathfrak{e} \quad Cd \end{array}$$

[26] and [27] suggest that a more general rule can be formulated to cover [25], [26] and [27]:

$$[25/26/27*] \quad \begin{array}{cc} M & M \\ / \backslash & / | \\ O & \mathfrak{e} \quad Cd \\ [-hi] & \\ [-lo] & \end{array} \longrightarrow \begin{array}{cc} M & M \\ / \backslash & / \backslash \\ O & V \quad Cd \\ [-hi] & \\ [-lo] & \end{array}$$

The rule, however, runs into trouble when the resultant sequence of mid vowel and coda constitutes an illformed rime, e.g. $\epsilon:j$. [28*] is a measure to avoid this. To ensure the simplicity of [25/26/27*] and [28*], the latter can be ordered after the former. Thus:

$$[28*] \quad \begin{array}{cc} M & M \\ | \backslash & / | \\ O & \epsilon \quad j \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / \\ O & \epsilon \end{array}$$

[29] suggests that \mathfrak{e} and the mid V can swop positions in the input form without altering the output. [30] further suggests that \mathfrak{e} can be replaced by the more general [+low] vowel. the examples in [29] and [30], however, are too isolated to establish the regularities just mentioned.

[31] represents yet another mode of coercion.

$$[31*] \quad \begin{array}{cc} M & M \\ / \backslash & / | \\ O & i \quad e \quad Cd \\ [+cor] & \\ [-ant] & \end{array} \longrightarrow \begin{array}{cc} M & M \\ / \backslash & | \\ O & e \quad Cd \\ [+cor] & \\ [-ant] & \end{array}$$

[32] represents another case of rime deletion, this time before a uniform (i.e. unchecked) rime. [33] requires that the deleted V in [32] must be [-round]. Thus:

$$[32*] \quad \begin{array}{cc} M & M \\ / \backslash & | \\ O & V_1 \quad V_2 \\ [-rou] & \end{array} \longrightarrow \begin{array}{cc} M & M \\ | \backslash & / \\ O & V_2 \end{array}$$

[34] has the first M shared by a velar stop and a following rounded V. Perceptually at least, this resembles the labiovelar gw- or kw-. It is hard to say whether coercion has actually taken place.

A by-product of coercion is the elimination of vowel clusters. In this connection, [36] likewise gives rise to the elimination of consonant clusters. Both [36] and the various coercion modifications, then, actualize Lass' CS characteristic (f) cited in Section 10.1.0 above.

[20] to [34] do not include any example with trochaic (long V, short Cd) rime in the second syllable of the IC form. Such examples are very rare. Consider the following:

[37] GLOSS	IC	C O N T R A C T E D	
		Plain	Coerced
university	da:j ⁶ hɔ:k ⁶	da ⁶ (h)ɔ:k ⁶	dɔ:k ⁶
Yes!	hɛj ⁶ a:k ³	hɛ ⁶ a:k ³	ha:k ⁶
OK!	how ² a:k ³	ho ² a:k ³	—
would rather	tsej ⁴ jy:n ²	tse ⁴ jy:n ²	tsy:n ⁴⁺²

The coerced forms seem to suggest that trochaic rimes work in the same way as uniform (i.e. unchecked) rimes. But an examination of the four plain contracted forms reveals that only the first syllable is surely contracted. As such the plain contracted forms here belong with those in [14]. In [37] coercion of the type found in [32] does take place, with a similar constraint too, namely that the deleted V must be unrounded. But unlike that in [32], coercion in [37] involves the deletion of a mora, the only mora remaining for the first syllable. This compares with the deletion involved in Cl- formation. At any rate, [37] in fact represents a different kind of modification from [20] to [33].

Apart from [38], there are other forms of bisyllabic contractions that are even more isolated:

[38]	IC	CONTRACTED
"how many"	gej ² dɔ:l	ge ² ə:l / gɛ:2+1
"please"	([ɱ] ⁴)gɔ:j ¹ lej ⁵	gɔ:j ¹⁺⁵
"the next"	dɛj ⁶ ji:6	dɛj ⁶

We shall not bother to account for these isolated modifications, partly because of their isolatedness and partly because the contracted forms are more or less lexicalized.

The contraction of sep^6 "ten" to $[\text{e}:]^6$ is worth special mention. The contraction is peculiar in that both the onset and the coda are deleted in the monomoraic form. Granted the e: form is already lexicalized, the double deletion is nevertheless not unmotivated: it is bounded by # neither in front nor behind in its usual context, e.g. in $\text{sej}^3 \text{sep}/[\text{e}:]^6 \text{jet}^1$ "forty-one". Recognizing the lexicalized e:^6 and against the background of this section, we can readily account for the form sa:^{1+6} "thirty", which is often erroneously represented as sa:^1 (e.g. Ráo et al 1981:186). Consider the following array:

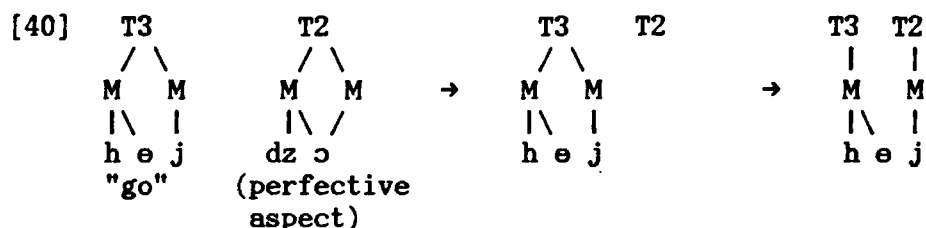
[39]	GLOSS	IC	CONTRACTED	
			Plain	Occluded
	twenty	$\text{ji:}^6 \text{e:}^6$	$\text{ji}^6 \text{e:}^6$	ja:^6
	thirty	$\text{sa:m}^1 \text{e:}^6$	$\text{sa}^1 \text{e:}^6$	sa:^{1+6}
	forty	$\text{sej}^3 \text{e:}^6$	$\text{se}^3 \text{e:}^6$	se:^{3+6}
	fifty	$[\eta]^5 \text{e:}^6$	<i>(Not applicable)</i>	
	sixty	$\text{lok}^6 \text{e:}^6$	$\text{lo}^6 \text{e:}^6$	lo:^6
	seventy	$\text{tset}^1 \text{e:}^6$	$\text{tse}^1 \text{e:}^6$	tsa:^{1+6}
	eighty	$\text{ba:t}^3 \text{e:}^6$	$\text{ba}^3 \text{e:}^6$	ba:^{3+6}
	ninety	$\text{gew}^2 \text{e:}^6$	$\text{ge}^2 \text{e:}^6$	ga:^{2+6}

As the array shows, sa:^{1+6} is just one of a series of coerced forms

¹ Cf. the contracted form dej^6 (lengthened) cited in Yuán et al 1960:194.

that result from the contraction of a numeral plus the following ɐ:6.

Again against the background of this section, the phenomena of tone stability described in Section 3.2.1 can be viewed as just one particular type of bisyllabic fusion, one in which everything in the second syllable except the tone is lost. The unassociated, or "floating", tone is then reassociated with the second M of the preceding syllable (which M is at the same time dissociated from the tone of the first syllable). For example:



As we have noted in Section 3.2.1, the conditions for the application of this kind of process are not phonological but lexical/morphological. We shall therefore not bother to go through these conditions. Suffice it to say that the mechanism developed in this section, especially principle [18], makes possible a simple and coherent account of the modifications involved.

10.2 Intonation

The most salient feature of Cantonese intonation is ironically that the language has little intonation. Two typological features of Cantonese go hand in hand with this scarcity of intonation. One has to do with lexical tone, as can be expected. Not only is Cantonese a tone language, but it has one of the richest tonal systems in the world. And not only is the number of contrastive tones in Cantonese one of the greatest, but the tonal system exploits both pitch height and pitch orientation at the same time. There are tone languages which primarily exploit pitch-height differences, like most African tone languages, and those that primarily exploit pitch-orientation differences, like Mandarin. Either of these two types of tone language leaves considerable room for the intonational exploitation of pitch. Cantonese, on the other hand, exploits both height and orientation for tonal contrasts, leaving little room for the intonational manipulation of pitch.

The other feature is that Cantonese has a rich system of sentence(-final) particles (SP). Yau (1980:35) records 206 such particles, built up from 42 free and 11 bound monosyllabic SPs. As Yau (p.35) notes:

In Cantonese (...) sentential connotations are mainly conveyed by means of a segmental device, the use of sentence particles. To obtain a certain intended sentential connotation an appropriate sentence particle has to be chosen and attached to the end of the sentence (...).

As such, SPs fulfil more or less the same function as intonation.¹ Yau (p.15) makes further inference from this functional overlap:

There is evidence supporting a tentative implication universal that there is a mutual compensation between SP and intonation patterns and that the more a language relies on the use of SP in expressing sentential connotations, the less significant will be the role played by intonation patterns, and vice versa.

The extent to which tone and SP are correlated is not clear, nor is the exact implicational relationship between tone, SP and intonation. But the fact that the three things are interrelated is beyond doubt: tone shares the form and SP shares the content of intonation.

Despite the traditional dichotomy of intonation language and tone language (e.g. in Abercrombie 1967:104), tone and intonation need not be mutually exclusive. Thus Cruttenden (1986:10-1) writes:

Tone and intonation are not completely mutually exclusive in languages. Languages with tonal contrasts may nevertheless make use of a limited amount of superimposed intonation. Such superimposed intonation may be manifested in four different ways:
(i) the pitch level of the whole utterance may be raised or lowered;
(ii) there will usually be downdrift in the absolute value of tones

¹ This mainly applies to the demarcative and attitudinal functions of intonation and Cantonese SPs. The accentual function of intonation is notably not the province of Cantonese SPs, for which function Cantonese resorts to intensity and duration differences.

- but downdrift may be suspended;
- (iii) the range of pitch used may be narrower or wider;
- (iv) the final tone of the utterance may be modified in various ways.

All these four ways are exploited in Cantonese, but to different extents. Section 10.2.1 considers the first three one by one, followed by Section 10.2.2 which is devoted to the fourth alone.

10.2.1 Intonation in the narrow sense

Cruttenden (p.9) describes intonation as "recurring pitch patterns, each of which is used with a set of relatively consistent meanings, either on single words or on groups of words of varying length." As such intonation has intonation-groups as its locus. That is to say its realization is not localized. It follows that only Cruttenden's (i), (ii) and (iii) are intonation in its proper sense. We consider these three one by one.

Consider (a) the already great pitch-height difference between T1, T2 at the high end and T4 at the low end, and (b) the relevance of pitch-level manipulation at the level of utterance. It should be clear that there is not much room for (i), i.e. pitch-level manipulation at the level of utterance. The extent to which one can overcome the difficulty imposed by (a) depends on the one hand upon the natural pitch range of the speaker, and on the other upon how far the speaker manages to compress the tonal pitch range, which compression is itself an intonation device and will be discussed later. Pitch range compression, however, makes the difficulty imposed by (b) even graver. The strategies available to overcome this latter difficulty include using the contoured variant of tones (this applies to T1 and T4) and signalling the manipulation of pitch level of utterance by other means. In connection with this second strategy, raised pitch level is sometimes accompanied by falsetto voice and marked facial expression, and lowered pitch level by creaky voice and suppressed level of movement of the vocal organs, especially the lips and the lower jaw. At any rate the extent to which pitch level of utterance is manipulated intonationally

varies greatly from person to person.

For those who do exploit pitch level of utterance, raised pitch level may signify emphasis, surprise, etc. With fast tempo it may suggest a critical moment. Accompanied with caricaturing devices (in terms of voice quality, facial expression, particular setting of the vocal organs, etc.) it may signify caricatured citation of direct speech. Lowered pitch level, on the other hand, is typical of grumbling but may also suggest hesitation and/or lack of confidence. The uncertainty of signification suggested here can be eliminated by the appropriate use of SP.

Cruttenden's (ii) involves two things: downdrift, and the suspension of downdrift. The latter presupposes the former. Downdrift, or declination, "refers to the fact that the pitch of the voice is most commonly lower at the end of a sentence than it is at the beginning." (Cruttenden 1986:126) This account of downdrift does not distinguish between (a) the gradual decline of the pitch level throughout the utterance, and (b) the fall in pitch at the very end of the utterance. This is unfortunate because the distinction is one we would like to draw, in view of the fact that (b) would be classed under Cruttenden's (iv), to be discussed in the next section. I am using "downdrift" to refer to (a) only.

Note that (a) implies (b) (but not vice versa). Thus, when Vance (1976:389) says that the results of his experiment "indicate that in Cantonese (...) final tone lowering occurs in ordinary declarative sentences," it is not clear if this manifests (a) or (b). His data in detail, however, show that the purely even tones, i.e. T3 and T6, have a strong tendency to be realized as a fall, no matter whether the tone occupies a medial or a final position in the sentence. Another revealing fact also escaped the attention of Vance: the sentence-initial word kej⁵ "he" is constantly of higher pitch than the sentence-medial word si:⁵ "city" (which is one of the six test words used in the experiment). But it is only natural that he was not aware of this interesting fact, for he actually thought that kej⁵ was in T2, which is precisely the higher-pitched counterpart of T5. To conclude, final tone lowering in Vance's experiment manifests not (b) but (a).

Vance's experiment covers declarative sentences only.

Impressionistically downdrift can be detected in unmarked utterances in general. Downdrift, however, does not apply when monosyllables are cited in isolation. If suspension of downdrift is recognized at all, here is where it applies. It is interesting that bisyllables, even when cited for illustration, often trigger downdrift. In a sense, then, speakers tend to view multisyllabic utterances as ordinary discourse rather than citation.

Cruttenden's (iii) concerns the elasticity of pitch range. I know of no study on this aspect of Cantonese intonation, except for the brief assertion "Intonation in Cantonese may manifest itself in a reduction or exaggeration of the contour of the lexical tones, or even temporarily obliterate their contrasts." (Kwok and Luke 1983), but elasticity of pitch range is doubtless available for exploitation. Thus, in general wider range indicates excitement and suggests vividness of expression.¹ Narrower range by itself indicates just the opposites. However, when compression of pitch range paves the way for and is accompanied by the kind of pitch-level manipulation discussed above, the effects of the latter manipulation take over.

10.2.2 Modification of the final tone

Despite the fact that it may not be the proper kind of intonation linguists usually speak of, the importance of final tone modification in Cantonese intonation / sentential connotation cannot be over-emphasized. Kwok and Luke 1983 and 1985, the only general descriptions of Cantonese intonation I know of, are a little misleading in that the authors speak of Cantonese intonational contrasts in terms that are usual for intonation languages, e.g. English. On close examination, however, we find that unlike in English the realization of such intonational contrasts are all localized: they are all utterance-final phenomena. The contrasts recognized in Kwok and Luke 1985 are summarized below:

¹ Professional story telling typically makes use of a larger-than-usual pitch range.

[41]

FORM	OTHER MODIFICATIONS	FUNCTION	EXAMPLE
+rise		(a) Change state- ments to questions	lej ⁵ wu:j ⁵ həj ³ "You will go."
ditto	Shortened	(b) Change ques- tions to echo- questions	kəj ⁵ dej ⁶ ge:j ² si: ⁴ fa:n ¹ "When will they be back?"
ditto		(c) Soften requests	si: ³ ha: ⁵ jon ⁶ la:m ⁴ sek ¹ "Try blue."
+ abrupt rise	Higher pitch level of utterance, louder	(d) = (a) plus surprise	
+fall		(e) Harden requests	ha:ŋ ⁴ gwo: ³ lej ⁴ "Come over here."
+ risefall (sharp)	Lengthened	(f) Contrastive stress	jəw ⁵ "Yes."
+ risefall (blunt)	Lengthened	(g) Irony	ŋo: ⁵ [m] ⁴ fa:n ⁴ "I am not upset."

We said earlier that intonation in Cantonese shares its form with tone and its content with sentence particles. When Kwok and Luke discuss Cantonese intonation, insufficient consideration is given to tone and SP. In any discussion of Cantonese intonation, the question must not be dodged as to how intonation faces competition from tone for the signifier, i.e. pitch-shape, and from SP for the signified, i.e. sentential connotations.

Consider (c) and (e) in [41]. If the respective intonation contrasts are valid, the examples are badly chosen. In the absence of other examples, how do we know that [+rise] for (c) is not due to the final T1, the highest tone, and that [+fall] for (e) is not due to the final T4, the lowest tone? The reason given for the lack of control sentences is that wording restricts the connotational possibilities. But if this is true, how do we know that the correspondence extracted between form and content is useful at all? Moreover, the denotation of the two sentences under discussion can easily be overridden by the connotations given by

These two SPs are usually lumped together and assigned T3. As we have shown in the notation of [43], a distinction can and should be made between T3 and the pitch shape of the second "la:".

This is not to deny that some of the pitch shapes of SP resemble one or another of the eight tone-shapes of the six basic tones. I assigned T3 to the first "la:" above, for instance. But I claim that the tones so identified for SP result from coercion, a recurrent concept in this thesis. Thus, SP pitch shapes are not confined by tones, though they can be coerced into tones.

There are SPs that have intrinsic pitch shape but no intrinsic segments. Consider the following examples:

[44] $\left. \begin{array}{l} \text{si:1+H} \\ \text{si:2+H} \\ \text{si:3+H} \\ \text{si:4+H} \\ \text{si:5+H} \\ \text{si:6+H} \end{array} \right\} \text{ "Was it } \left\{ \begin{array}{l} \text{si:1} \\ \text{si:2} \\ \text{si:3} \\ \text{si:4} \\ \text{si:5} \\ \text{si:6} \end{array} \right\} \text{ you just said?"}$

Echo questions can be formed by attaching a high-pitched segmentless tail to any utterance (which may be of whatever sentence types: statemant, question, command, etc.) of any length. Not only can the resultant tone be distinguished from T3-T6, which do not usually end high, but it can be distinguished from T1 and T2 as well by virtue of peculiar contour, extra length and possibly extra hight of pitch. The descriptive device of autosegmental phonology and the idea of autosegmental morae enable us to account for the phenomenon by

H
representing the SP in question as $\left| \begin{array}{c} \text{H} \\ \text{M} \end{array} \right|$, which is not pre-linked to any
M

segment. Subsequent association of this segmentless SP with the preceding syllable accounts for the extra length and complex "tone" of the syllable.

All SPs are enclitics. No #, therefore, lies between the final syllable and the attached SP. It follows that contraction in the form of mora deletion is possible, which results in a syllable of ordinary length (i.e.

bimoraic), but nevertheless with a complex tone, i.e. its original tone plus the SP pitch shape. In general, following the contraction, the maximally identifiable sequence of tone + SP-pitch may undergo simplification in the form of smoothing or corner-cutting, so that the recoverability of the tone weakens. With the possibility of such simplification fully recognized, the following remark by Kwok and Luke (1983) will not be incompatible with my account:

[L]exical tones in Cantonese, instead of invariably retaining their canonical forms in connected speech, are more often than not drastically reshaped by intonation, so much so that for instance, a low-fall may become a sharp rise, and a high-rise may become a rise-fall.

If reshaping by intonation is in the form of reshaping by segmentless SP as outlined above, then such reshaping does not necessarily mean the irrecoverability of the tone affected. Even in cases where the tone *is* irrecoverable, the irrecoverability can be readily explained by neutralization due to smoothing/corner-cutting which is in turn due to (optional) contraction. Smoothing/corner-cutting predicts that it is the ending point rather than the starting point of a tone shape that is modified. This prediction is also corroborated by Kwok and Luke's (1983) observation:¹

[I]t seems that the distinction of the three pitch levels of high, mid and low is retained in that the pitch level at which the last syllable in a tone-group *starts* is still governed by whether the canonical form of its lexical tone belongs to the high, mid, or low category.

Note that exactly three pitch heights are recognized at the starting point of tones, namely high (for T1), mid (for T3) and low (for T6, and redundantly for T2, T4 and T5). In the light of the foregoing discussion, the intonation contrasts in [41] can all be given SP interpretations. Thus, I claim that the slightly rising fa:n¹ in (b) results from the contraction of fa:n¹ plus the echo-question SP.² (a)

¹ The emphasis is mine.

² Kwok and Luke claim that the syllable in the echo question is 'shortened'. Their F₀ graphs, however, show that the shortening is

and should also be analysed likewise. (a) as an ordinary question is odd. As Vance (1976:376) notes:

Yes-no questions in Cantonese do not have a final rise in pitch. Such questions are formed with an overt interrogative particle, and Cantonese-English bilinguals brand attempts to impose such a rising contour as corruptions from English.

The only natural interpretation of (a) as a question is that it is an echo question as (b) is. The final item in (a), i.e. hej³, combines with the segmentless echo-question SP to give a rising pitch-shape. Even (d) need not be analyzed differently: it can be viewed as just another instance of echo question. The abrupt rise can be explained by quickened tempo. Quickened tempo, rise in utterance-wide pitch level and increase in loudness are all correlates of surprise in their own right.

The highly productive device of attaching a rise-fall tail to an utterance to signal irony, as illustrated by [41(g)], and the extra length of the ironical sentence compared with the plain sentence, mean that an

RF
irony SP of the form | can be established. The following examples show
M

how the tone of the final syllable can be retained despite the syllable's burden of realizing the SP.

[45]	ŋɔ:5 [m]4 sem1	$\left\{ \begin{array}{l} \text{gɛp}^{1+\text{RF}} \\ \text{sej}^{2+\text{RF}} \\ \text{toŋ}^{3+\text{RF}} \\ \text{fa:n}^{4+\text{RF}} \\ \text{leŋ}^{5+\text{RF}} \\ \text{ly:n}^{6+\text{RF}} \end{array} \right\}$	"I am not	$\left\{ \begin{array}{l} \text{impatient."} \\ \text{disillusioned."} \\ \text{distressed."} \\ \text{perturbed."} \\ \text{grateful."} \\ \text{upset."} \end{array} \right\}$
------	----------------	--	-----------	--

The form of the SP in (f) is very similar to that in (g). Although (f) is recognized as "sharp" to be distinguished from "blunt" for (g), I feel that they are merely different manifestations of the same emphatic/irony

minimal, and is more likely accidental than regular. But at least it should be clear that the echoed fa:n¹ is not longer than the original fa:n¹, suggesting that contraction has taken place.

[SP. This is especially so because a single-item (in fact monosyllabic)]

sentence is a strange example to use to illustrate contrastive stress: contrastive stress in such a sentence will not be distinguishable from mere emphasis. I am, however, not committed to such details. What is more relevant is the fact that no matter whether it is collapsible with RF(blunt) or not, RF(sharp) can likewise be viewed as a segmentless SP.

Only (c) and (e) are not yet discussed. These two differ from the others in that so far as pitch-shape is concerned, the intonation form for (c) resembles the synonymous SP ´ə in [42] and the intonation form in (e) resembles the synonymous `a: in [42e]. Apart from treating the intonation forms as segmentless SPs, as we have been doing so far, the possibility is open for us to account for the final tonal modification in terms of segmental SP, followed by segment deletion together with tone stability. Again I am not committed to a particular treatment. Either solution follows from the list of claims given earlier, which is the main concern of this section.

Accepting the account of Cantonese intonation developed above, the problem of the SP-intonation-tone relation can be tackled in fresh light. The tension between SP and intonation is resolved by recognizing that the bulk of Cantonese intonation is in the form of final tone modification, which in turn is derived from SP. The tension between intonation and tone is resolved by recognizing that final tones can be retained after modification, though subject to optional neutralization.

R E F E R E N C E S

- Abercrombie, D 1967. *Elements of general phonetics*. Edinburgh: Edinburgh U Press.
- Anderson, S R 1985. *Phonology in the twentieth century*. Chicago: The U of Chicago Press.
- Asher R E and E J A Henderson (eds) 1981. *Towards a history of phonetics*. Edinburgh: Edinburgh U Press.
- Bái, Wǎnrú 1982. "Guǎngzhōu-fāngyán liándú yīn-biàn jǔlì (= Examples of pronunciation modification in Cantonese connected speech)" *Fāng-yán*. 1982:19-21.
- Ball, F D 1899-900. "Prefatory Remarks" to Ch'an 1899-90.
- Barnett, K M A 1949-50. "A transcription for Cantonese: notes on Mr. Yuan Ren Chao's *Cantonese Primer*" *Bulletin of the SOAS*. Xiii, Pt 3: 725-45.
- Bauer, R S 1979. "Alveolarization in Cantonese: a case of lexical diffusion" *J of Chinese Linguistics*. 7:132-43.
- 1982. *Cantonese sociolinguistic patterns: correlating social characteristics of speakers with phonological variables in Hong Kong Cantonese*. Ann Arbor: U Microfilms International.
- 1983. "Cantonese sound change across subgroups of the Hong Kong speech community" *J of Chinese Linguistics*. 11:301-54.
- 1984. "The expanding syllabary of Hong Kong Cantonese". MS.
- 1986. "The microhistory of a sound change in progress in Hong Kong Cantonese" *J of Chinese Linguistics*. 14:1-42.
- Bazell, C E et al 1966. *In memory of J. R. Firth*. London: Longman.
- Bloomfield, L 1933. *Language*. London: George Allen & Unwin.
- Brosnahan, L F and B Malmberg 1970. *Introduction to phonetics*. Cambridge: Cambridge U Press.
- Canepari, L 1983. *Phonetic notation*. Venice: Università di Venezia.
- Catford, J C 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh U press.
- Cén, Qíxiáng 1946. "Guǎngzhōu-yīnxì gàishuō (= Outline of the phonetic system of Cantonese)" *Guǎngdōng Jiànshè*. 1-2:1-9.
- 1982. *Guójī yīnbīāo* (= The International Phonetic Alphabet). Húběi: Húběi Rénmín.
- Ch'an, C-S K H 1899-900. "Rules for the use of variant tones in Cantonese" *China Review*. XXIV:209-26.

- Chao, Y-R 1930. "A system of tone letters" *lǎ mǎ:trə fonetik*. XL:24-7.
- 1931. "fǎnqiè-yǐ bā zhǒng (=Eight varieties of secret language based on the principle of fǎnqiè)" *Bulletin of the Institute of History and philology*. 2:320-54.
- 1947. *Cantonese Primer*. Cambridge, Mass.: Harvard U Press.
- Chén, Hùiyīng and Bái, Wǎnrú 1958. "Guǎngzhōu-yīn hé Běijīng-yīn de bǐjiào (= Comparing Cantonese and Mandarin sounds" in Chén et al 1958.
- Chén, Hùiyīng et al 1958. *fāngyán hé Pǔtōnghuàcóngkān (= Collected Essays on dialects and Putonghua)*. Běijīng:Zhōnghuá.
- Chen, M Y 1984-5. "From Middle Chinese to Modern Cantonese" *J of Chinese Linguistics*. 12:148-98, 334-88; 13:122-70.
- Cheng, Teresa 1968. *The phonological system of Cantonese*. Berkeley: Project on Linguistic Analysis. 2-5-1.
- Cheng, C C 1973. *A synchronic phonology of Mandarin Chinese*. The Hague: Mouton.
- Cheung, H-N S 1972. *Studies on Cantonese as spoken in Hong Kong*. Hong Kong: The Chinese U of Hong Kong.
- Cheung, Y-S 1969. "Xiānggǎng Yuèyǔ yīnpíng-diào jí biàndiào wèntí (= A study on the upper even tone and tone sandhi in the Cantonese dialect as spoken in Hong Kong)" *The J of the Institute of Chinese Studies of the Chinese University of Hong Kong*.
- 1986. "Xiānggǎng Guǎngzhōu-huà Yīngyǔ-yīnyì jiècí de shēngdiào gūizé (= On the tone system of loan words from English in Hong Kong Cantonese)" *Zhōngguó Yǔwén*. 1986-1:42-50.
- Chiang, K C (n.d.; c.1940). *A practical English-Chinese Dictionary*. Singapore: Lam Yeong Press.
- Ching, Y-C T 1981. "Communication of lexical tone patterns in Cantonese" PhD thesis, U of London.
- Chomsky, N 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.
- Chomsky, N and M Halle 1968. *The sound pattern of English*. New York: Harper & Row.
- Chou, F-K 1968. "Lùn Qiè-yùn yīn (= On the sound pattern of Middle Chinese)" reprinted in Chou 1984.
- 1984. *Zhōngguó yīnyùnxué lùnwén-jī (= Collected Essays on historical Chinese phonology)*. Hong Kong: The Chinese U of Hong Kong Press.
- Clements, G N and K Ford 1979. "Kikuyu tone shift and its synchronic consequences" *Linguistic Inquiry*. 10:179-210.

- Clumeck, H et al 1981. "The aspiration contrast in Cantonese word-initial stops: data from children and adults" *J of Chinese Linguistics*. 9:211-25.
- Cohen, A and M P R Van den Broeke (eds) 1983. *Abstracts of the Tenth International Congress of Phonetic Sciences*. Dordrecht:Foris.
- Cruttenden, A 1986. *Intonation*. Cambridge: Cambridge U Press.
- Dell *
Dow, D-M F 1972. *An introduction to the pronunciation of Chinese*. Britain: F D-M Dow.
- Egerod, S 1956. *The Lungtu Dialect*. Copengagen: Efnar Mundsgaard.
- Eitel, E G 1887. *A Chinese dictionary in the Cantonese dialect*. London: Trübner.
- Fàn, Guó 1979. "Yuèyǔ-kǒuyǔ zhōng míngcí-xìng diécí zhī biàndiào gūilǚ (= Pattern of tonal modification with respect to nominal reduplication in colloquial Cantonese)" *Yǔwén Zāzhǐ*. 2:39-44.
- Féng, Siyǔ 1962. *Guǎngzhōu-yīn zǐhùì (= A Cantonese glossary)*. Hong Kong: Shìjiè.
- Firth, J R 1948. "Sounds and prosodies" reprinted in Palmer 1970:1-26.
- Fok, C Y-Y 1974. *A perceptual study of tones in Cantonese*. Hong Kong: U of Hong Kong.
- Fung, T 1974. *Yuèyǔ tóngyīn zǐdiǎn (= A dictionary of homophones in Cantonese)*. Hong Kong: Dōng Lián.
- Dell, F 1980. *Generative Phonology*. Cambridge: Cambridge U Press.
- Fudge, E C 1986. "Branching structure within the syllable" Handout, LAGB Spring Meeting 1986.
- Gāo, Huáníán 1980. *Guǎngzhōu-fāngyán yánjiū (= Studies on the Guǎngzhōu dialect)*. Hong Kong: Commercial Press.
- Gandour, J 1981. "Perceptual dimensions of tone: evidence from Cantonese" *J of Chinese Linguistics*. 9:20-36.
- Goldsmith, J A 1979. *Autosegmental phonology*. London:Garland.
- Halle, M and G N Clements 1983. *Problem book in phonology*. London: The MIT Press.
- Halle, M and K N S Stevens 1979. "Some reflections on the theoretical bases of phonetics" in Lindblom and Ohman 1979:335-49.
- Halliday, M A K 1981. "The origin and early development of Chinese phonological theory" in Asher and Henderson (eds) 1981:123-40.
- Haraguchi, S 1977. *The tone pattern of Japanese: an autosegmental theory of tonology*. Tokyo:Kaitakusha.

Harris, J 1985. *Phonological variation and change: studies in Hiberno-English*. Cambridge: Cambridge U Press.

- [Harris, Z S 1944. "Simultaneous components in phonology" *Language*. 20: 181-205.
- 1986. "The lexicon in phonological variation" in *Perspectives on the English language in Ireland* (Dublin: Trinity College).

Hashimoto, M and O-K Y Hashimoto 1968. "Phonological redundancy and Cantonese vowels" *Gengo Kenkyu*. 53:101-11.

Hashimoto, O-K Y 1971. "A note on D. Jones & K.T. Woo's A Cantonese phonetic reader" *Onsei no Kenkyu*. 15:79-94.

— 1972. *Phonology of Cantonese*. Cambridge: Cambridge U Press.

Hill, T 1966. "The technique of Prosodic Analysis" in Bazell et al (eds) 1966:198-226.

Hockett, C F 1954. "Two models of grammatical description" *Word*. 10: 201-34.

— 1955. *A manual of phonology*. Baltimore: Waverly Press.

— 1958. *A course in modern linguistics*. New York: Macmillan.

Huang, R 1965. *English pronunciation explained with diagrams*. Hong Kong: Hong Kong U Press.

Hudson, R 1984. *Invitation to linguistics*. Oxford: Martin Robertson.

Hulst, H v d and N Smith 1982 "An overview of Autosegmental and Metrical Phonology" in Hulst and Smith (eds) 1982:1-45.

— (eds) 1982. *The structure of phonological representations (Part I)*. Dordrecht:Foris.

— 1984. "The framework of Nonlinear Generative Phonology" in Hulst and Smith (eds) 1984:1-55.

— (eds) 1984. *Advances in Nonlinear Phonology*. Dordrecht:Foris.

Hyman, L M 1975. *Phonology: theory and analysis*. London: Holt, Rinehart and Winston.

The International Phonetic Association 1949. *The principles of the International Phonetic Association*. London: the IPA.

Iwata, R 1985. "Syllable final stops in Southern Chinese" *Gengo Kenkyu*. 87:22-39.

Jakobson, R 1971 (Transl. by B. Hrushovski 1980). "Sign and system of language: a reassessment of Saussure's doctrine" reprinted in Jakobson 1985.

— 1985. *Verbal art, verbal sign, verbal time*. Oxford: Basil Blackwell.

- Jakobson, R and M Halle 1968. "Phonology in relation to phonetics" in B Malmberg (ed) 1968.
- Jakobson, R et al 1952. *Preliminaries to speech analysis*. Cambridge, Mass.: MIT Press.
- Jones, D 1912. "Introduction" to Jones and Woo 1912.
- 1957. "The history and meaning of the term 'phoneme'" in Jones and Laver (eds) 1973.
- Jones, D and K T Woo 1912. *A Cantonese phonetic reader*. London: U of London Press.
- Jones, W E and J Laver 1973. *Phonetics in linguistics: a book of readings*. London: Longman.
- Kam, T H 1977. "Derivation by tone change in Cantonese" *J of Chinese Linguistics*. 5:186-210.
- Kao, D L 1971. *Structure of the syllable in Cantonese*. The Hague: Mouton.
- Karlgren, B 1915-26. *Études sur la Phonologie chinoise*. Stockholm: Gotembourg.
- 1923. *Analytical dictionary of Chinese and Sino-Japanese*. New York: Dover (1974).
- Katz, J J (ed) 1985. *The philosophy of linguistics*. Oxford: Oxford U Press.
- Killingley, S Y 1982. *The grammatical hierarchy of Malayan Cantonese*. Newcastle Upon Tyne: Grevatt & Grevatt.
- 1985a. "A non-instrumental experiment on Cantonese phonological tone: five or six?" Handout, LAGB autumn meeting 1985.
- 1985b. *A new look at Cantonese tones: five or six?* Newcastle Upon Tyne: S Y Killingley.
- Kiu, K L 1977. "Tonal rules for English loan words in Cantonese" *J of the International Phonetic Association*. 7:17-22.
- Kratochvil, P 1968. *The Chinese language today*. London: Hutchinson.
- Kwok, H and K K Luke 1983. "Intonation in Cantonese" in Cohen and Broecke (eds) 1983.
- 1985. "Pǔtōnghuà yǔ Yuè-yǔ yǐdiào de chūbù bǐjiào (= A preliminary comparison of Pǔtōnghuà and Cantonese intonation)" Handout, Seminar on Pǔtōnghuà teaching and evaluation, 1985, Hong Kong.
- Lau, S 1973. *A Cantonese-English and English-Cantonese glossary to accompany Elementary Cantonese (Lessons 1-20)*. Hong Kong: Hong Kong Government Printer.

- Lass, R 1984. *Phonology: an introduction to basic concepts*. Cambridge: Cambridge U Press.
- Lee, H-T T 1985. "Guǎngzhōu-huà yuányīn zhī yīnzhì jí cháng-duǎn dùilì (= The phonetic quality and long/short distinction of Cantonese vowels)" *Fāngyán*. 1985-1:28-38.
- Lǐ, Róng 1954. *Qiè-yùn yīnxǐ* (= *The phonological system of Qiè-yùn*). Běijīng: Kēxué.
- Light, T 1977. "The Cantonese final: an exercise in indigenous analysis" *J of Chinese Linguistics*. 5:75-102.
- Linguistic bibliography. 1949 onwards. Utrecht: Spectrum.
- Lisker, L and A S Abramson 1964. "A cross-language study of voicing in initial stops: acoustic measurement" *Word*. 20:384-422.
- Lockhart, J H S 1882. "Canton Syllabary" *The China Review*. X:312-326.
- Lucas, A 1985. *Chinese linguistics: a bibliography: 1975-1982*. Paris: Langages croisés.
- Luke, Kang-Kwong 1983. "Xiānggǎng Yuèyǐ lǐ [u:] [y:] de yīnwèi wèntí (= On the phonemic status of [u:] and [y:] in Hong Kong Cantonese)" *Yǔwén Zāzhǐ*. 11:37-43.
- 1984. "Expedient and orientational language-mixing in Hong Kong" *York papers in linguistics*. 11:191-201.
- Lyovin, A 1973. Review of Kao 1971. *Language*. 49:954-61.
- Malmberg, B 1968. *Manual of Phonetics*. London: North-Holland.
- McCoy, W J 1966. "Szeyap data for a first approximation of proto-Cantonese" PhD Dissertation. Ithaca: Cornell U.
- Meetham, A R and R A Hudson (eds) 1969. *Encyclopaedia of linguistics, information and control*. Oxford: Pergamon Press.
- O'Connor, J D 1980. *Better English Pronunciation*. Cambridge: Cambridge U Press.
- Palmer, F R (ed) 1970. *Prosodic analysis*. London: Oxford U Press.
- Parker, E H 1880a. "Canton syllabary" *China Review*. VIII:363-82.
- 1880b. "Characterless Chinese words" *China Review*. IX:85-88.
- Phoon, M-K 1976. *A synchronic phonology of modern Cantonese dialect*. Singapore: Institute of humanities and social sciences, Nanyang U.
- Pike, K L 1947. *Phonemics*. Ann Arbor: The U of Michigan Press (1975).
- Rai, T 1958. "On the phonology of Cantonese" (article in Japanese) *Chūgoku Gogaku*. LXX-1:3-5.
- Ráo, Bǐngcái 1980. "Yuè-fāngyán zìyīn de dīngyīn wèntí (= On Cantonese

- Orthoepy)" *Yǔwén Zāzhǐ*. 5:42-5.
- Ráo, Bǐngcái et al 1981. *Guǎngzhōu-huà fāngyán cǐdiǎn* (= *Cantonese dialect dictionary*). Hong Kong: Commercial.
- Richards, J et al 1985. *Longman dictionary of applied linguistics*. Harlow: Longman.
- Saussure, F de 1922, *Cours de linguistique générale*.
- 1983. *Course in general linguistics* (Trans. by R Harris of Saussure 1922). London: Gerald Duckworth.
- Seers, O 1908. "Chinese phonetics: observations made on the sounds of the Cantonese dialect of Chinese" *lǎ mǎ:trə fōnetik*. 1908:30-4.
- Shàn, Zhōuyáo 1980. "Yuè-yīn yùn-hùì, Zhōngwén zǐdiǎn, Zhōnghuá xīn-zǐdiǎn zhōng yīxiē yǔ Xiānggǎng tōngxíng de shíjì dúyīn yǒu jùlǐ de Yuèyǔ zhùyīn = Certain pronunciations given in *Yuè-yīn yùn-hùì*, *Zhōngwén zǐdiǎn*, and *Zhōnghuá xīn-zǐdiǎn* which are at variance with the actual pronunciation current in Hong Kong" *Yǔwén Zāzhǐ*. 4: 46-8.
- Shimizu, S 1963-4. "Kantongo no /e/ (=The vowel /e/ in Cantonese)" (article in Japanese) *Studia Phonologica*. III:7-16.
- Simon W 1953. "Introduction" to Whitaker 1953.
- Sloat, C et al 1978. *Introduction to phonology*. Englewood Cliffs: Prentice-Hall.
- Stewart, A H 1976. *Graphic representation of models in linguistic theory*. London: Indiana U Pres.
- Stich, S 1972. Reprinted in Katz (ed) 1985:126-145.
- Toodoo, A 1957. *Chūgodugo on'inron* (= *Chinese phonology*) (in Japanese) Tokyo: Kōnan shoin.
- Trubetzkoy, N S 1931. "Phonologie et géographie lingistique" *Travaux du Cercle Linguistique de Prague*. 4:228-34.
- Tse, K-P 1978. "Tone acquisition in Cantonese: a longitudinal case study" *J of Child Language*. 5:191-204.
- Tse, S-M 1982. "The acquisition of Cantonese phonology" PhD thesis, The U of Columbia.
- Tsuji, N 1980. *Comparative phonology of Guangxi Yue dialects*. Tokyo: Kazama Shobo.
- Tung, T-H 1961. "Shēngmǔ yùnmǔ de guānniàn hé xiàndàide yǔyīn fēnxī lǐlùn (= The concept of 'initial' and 'final' and modern theories phonemic analysis)" *Bulletin of the Institute of History and Philology*. Spec. Pub. 4:681-91.
- 1964. "Bipartite division of syllables in Chinese phonology" in *Proceedings of the International Congress of Linguistics 9, 1962*.

- Vance, T J 1976. "An experimental investigation of tone and intonation in Cantonese" *Phonetica*. 33:368-92.
- 1977. "Tonal distinctions in Cantonese" *Phonetica*. 34:93-107.
- Vance, J T and C A Walker 1976. "Tone and intonation in Cantonese" *Canadian Linguistic Society*. 1976,12:640-655.
- Walton, A R 1983. *Tone, segment, and syllable in Chinese: a poly-dimensional approach to surface phonetic structure*. Ithaca: Cornell U.
- Wáng, Lǐ 1936-7. *Zhōngguó yīnyùn-xué* (= *Chinese phonology*). Shànghǎi: Shāngwù.
- 1985. *Hànyǔ yǔyīn shǐ* (= *History of Chinese sounds*). China: Zhōngguó Shěhuì Kēxué.
- Wang, W 1967. "Phonological features of tone" *Int'l J of American Linguistics*. 4:1-9.
- 1969. "Competing changes as a cause of residue" *Language*. 45:9-25.
- Wells, J C 1969. "Articulatory phonetics" in Meetham and Hudson 1969.
- 1982. *Accents of English*. 3 Vols. Cambridge: Cambridge U Press.
- Whitaker, K P K 1953. *1200 Chinese basic characters for students of Cantonese*. Londong: Lund Humphries.
- 1955-6. "A study on the modified tones in spoken Cantonese" *Asia Major*. 5:9-36; 184-207.
- Wong, M K S 1981. "A quantitative study of tone change in Cantonese: some preliminary findings" *Chicago linguistic society*. 1981,17: 419-30.
- 1982. "Tone change in Cantonese. PhD thesis, U of Illinois at Urbana-Champaign.
- Wong, S L 1940. *A Chinese syllabary pronounced according to the dialect of Canton*. Kong Kong: Chung Hwa (1974).
- Williams, S W 1856. *Tonic dictionary of Chinese language in the Canton dialect*. Canton: Chinese Repository.
- Wú, Zhìxūn 1976. "Lüè-tán Yuè-yīn dúfǎ chángshí" (= an outline of Cantonese orthoepy)" *New Asia College Chinese Department Annual* (Hong Kong). 10:1-24.
- Yang, S J P F-M 1974. *Chinese linguistics: a selected and classified bibliography*. Hong Kong: The Chinese U Press.
- 1981. *Chinese dialectology: a selected and classified bibliography*. Hong Kong: The Chinese U Press.
- Yau, S-C 1980. "Sentential connotations in Cantonese" *Fāngyán*. 1980,1:

- Yeung, S-W H 1980. "Some aspects of phonological variation in the Cantonese spoken in Hong Kong" MA dissertation. U of Hong Kong.
- Yip, M J W 1980. *The tonal phonology of Chinese*. M J W Yip.
- 1982. "Reduplication and C-V skeleta in Chinese secret languages" *Linguistic inquiry*. 13:637-61.
- Yú, Nǎiyǒng 1979. "Xiānggǎng Yuèyǔ yīngpíng-diào jí biàndiào de Yǎnbiàn gūilǚ yǔ lìshǐ xìngzhì (= The derivational pattern and historical characteristics of the yīngpíng tone and tone sandhi in Hong Kong Cantonese)" *Journal of Baptist College* (Hong Kong). 6:43-62.
- 1984. "Xiānggǎng guǎngzhōu-huà jǐzhǒng yǔ biànyì wúguān zhī zìyīn liǎngdú xiànxàng (= Some sub-lexical alternations in Hong Kong Cantonese)" *Yǔwén Zāzhì*. 12:60-72.
- Yuán, Jiāhuá et al 1960. *Hànyǔ fāngyán gàiyào (= Outline of Chinese dialects)*. Běijīng: Wénzì Gǎigé.
- Yǔyán Yánjiū Suǒ 1958. "Hànyǔ fāngyīn de jǐge wèntí (= Some phonetic problems of Chinese dialects)" in Chén et al 1958:149-74.
- 1978. *Zhōngguó yǔyánxué lùnwén suǒyǐn (= A bibliography of Chinese linguistics papers)*, Vol I. Běijīng:Shāngwù.
- 1983. *Zhōngguó yǔyánxué lùnwén suǒyǐn (= A bibliography of Chinese linguistics papers)*, Vol II. Běijīng:Shāngwù.
- Zēng, Zīfán 1982. *Guǎngzhōu-huà Pǔtōnghuà kǒuyǔ-cí duìyǐ shǒucè (= A manual of correspondences between Cantonese and Pǔtōnghuà colloquialisms)*. Hong Kong: Joint Publishing.
- Zhān, Bóhùì 1981. "Hànyǔ fāngyán yánjiū gàikwàng hé yǒugwān de jǐge wèntí = Chinese dialectology and related matters" *Yǔwén Zāzhì*. 8:3-11.
- Zhāng, Lìyán 1983. "yǒu-yīn wú-zì yīnjié zǒngjì (= Stocktaking characterless morpho-syllables in Cantonese)" *Yǔwén Zāzhì*. 10:28-35.
- Zōng, Fúbāng 1964. "guānyú Guǎngzhōu-huà yīnpíng-diào de fēnhuà wèntí (= Problems concerning the split of the yīnpíng tone in Cantonese)" *Zhōngguó Yǔwén*. 1964, 5:376-89.